OXFORD ECONOMIC PAPERS

NEW SERIES
VOLUME 25
1973

OXFORD
AT THE CLARENDON PRESS
1973

CONTENTS

VOLUME 25

Number 1, March 1973

Number 2, July 1973

Number 3, November 1973

ERRATA

Vol 25, No. 1, p. 94, Table 1: insert a final line under the heading *Allocation o gross profits.*

	Total	*Local*	*Foreign*	*Unallocated*
Dividends	24 4	9 1	15 3	—

Vol. 24, No. 3, p. 353, line 16: *for* 'shorter' *read* 'longer'.

Editorial Board

ALTHOUGH IT WAS originally intended as a channel for publication of articles b Oxford authors, contributions from elsewhere are welcome While most of th articles published will be concerned with Economics, the Editors also hope t accept articles on Economic History and Public Administration, provided the are likely to be of general interest to economists.

Books are not reviewed, but substantial review articles will be considere The Editors will also welcome bibliographical surveys designed to be of use t the general economist.

FACTOR MARKET DISTORTIONS, PRODUCTION, AND TRADE: A SURVEY

By STEPHEN P. MAGEE[1]

I. Introduction

THE traditional interest of economists in monopoly power and other distortions in product markets is shifting toward the problem of distortions and differentials in factor markets The purpose of this paper is to review a portion of the existing literature in this area and provide some minor extensions, with special emphasis on the implications for international trade. General studies which include at least brief discussions of the problem of factor market imperfections have been written by Bhagwati [16], Caves [31], Corden [34], Johnson [77], Linder [96], Myint [116], Taussig [149], and Viner [158] We do not consider here issues in the current controversy over the theory of capital (see Harcourt's survey [61]).

Several theoretical papers on the subject should be mentioned Four early works are by Cairnes [30], Manoilesco [104], Ohlin [118], and Viner [156] While Stolper and Samuelson [146] do not deal with distortions *per se*, their study of factor rewards in general equilibrium provides an important analytical tool in this area In the 1950s Eckaus [39], Haberler [52], Hagen [54], and Lewis [95] wrote on the subject while Bhagwati and Ramaswami [20], Bhagwati, Ramaswami, and Srinivasan [21], Fishlow and David [42], Johnson [77], Johnson [78], Johnson and Mieszkowski [80], and the present author [100] studied the problem in the 1960s. Since 1969 no fewer than thirty-six theoretical and empirical papers have appeared or are in draft form One of the most interesting early papers in the group is the qualitative study written by Ohlin [118] in 1931, in which he reviews Manoilesco's book [104]; it is interesting because it arrives at several of the conclusions which have been developed mathematically by the modern writers. A comprehensive modern treatment of the welfare effects is by Bhagwati [19], who has formulated a general theory encompassing the three causes and four possible types of economic distortions (factor distortions being only one of the four).

[1] This paper is a revision of Chapter 1 of the author's Ph D dissertation [100] He is indebted to the three members of his dissertation committee, Jagdish Bhagwati, Charles Kindleberger, and Paul Samuelson, to Robert E Baldwin, Charles Metcalf, J David Richardson, and other participants in the International Trade Workshop at the University of Wisconsin (where the paper was presented, Feb 1972), and to John Burton, Richard Caves, Robert Flanagan, Horst Herberg, Ronald Jones, Murray Kemp, Nan Magee, Carlos Rodriquez, and especially W M. Corden for discussions, suggestions, and comments He is grateful to the Woodrow Wilson National Fellowship Foundation, the Institute of International Studies, University of California, Berkeley, the National Science Foundation, and the Graduate School of Business, University of Chicago, for research support Naturally, no survey is comprehensive no doubt, some very important articles have been missed. Papers which came to my attention after the references were numbered have been included in footnotes

Several taxonomies have been suggested for classifying the causes and types of factor market imperfections. Bhagwati [19] cited three *causes* of differentials in his welfare paper: endogenous (due to some market imperfection under *laissez-faire*), and either autonomous or instrumental policy imposed differentials.

Following an earlier paper by Bhagwati [17], we can expand the traditional definition of a differential to incorporate two types. First, factor prices may be the same in all industries but there may exist a differential between real factor rewards and their marginal products in one or more industries. Second, real factor rewards may equal their respective marginal products in each industry but there may be a differential between the price of an identical factor in different industries. This expanded definition of a differential will thus incorporate more than one of the various inequalities which may arise in the factor market equilibrium conditions.

The word 'differential' is used here as a positive or purely descriptive term while 'distortion' denotes that a differential has certain normative or welfare implications (see Bhagwati and Ramaswami [20]). In the present context, a distortion can always be attributed to a differential of one sort or another, but not every differential implies that a distortion exists. Thus, a differential is a necessary but not a sufficient condition for a distortion.

For example, many studies have found factor price differentials which do not necessarily indicate distortion. The differentials may be caused by differences in

1. age and experience among workers in [1], [55], [59], [76], [91], [112],
2. education and skill reflecting a return on human capital in [1], [3], [20], [55], [59], [73], [75], [76], [83], [91], [112], [118], [123], [128], [136];
3. regional differentials due to moving costs or other factors in [1], [20], [25], [49], [58], [59], [105], [140], [162],
4. factor preference or disutility associated with particular industries or regions in [1], [42], [52], [68], [77], [118], [156], [157],
5. risk aversion in [41]; and
6. regional differences due to geographic concentration of low wage-low skill industries in [47], [58].

On the other hand, an even larger number of differentials have been cited as sources of welfare distortion. The causes of these differentials include

1. imperfect knowledge in [39], [42];
2. the rural-urban dichotomy in [54];
3. racism in [13], [27], [119], [150], [162],

4. monopoly power through unionism in [2], [20], [33], [78], [80], [94], [99], [110], [118], [125], [133], [136], [141], [151], [154],
5. monopoly power by one or both factors coupled with market power by producers in product markets (bilateral monopoly) in [2], [7], [26], [60], [61], [72], [130], [133], [137],
6. the maintenance and spread of union wage increases by 'pattern setting' in [98], [137];
7. seniority based on age or education which does not reflect economic superiority in [1], [76], [91], [162];
8. differences between the export- and import-competing sector's access to foreign capital in [36], or differences in ocean freight rates charged to the two sectors in [106];
9. discrimination against women or children in [17], [76], [91], [162];
10. collusion across industries by a factor which acts as a discriminating monopolist, charging different prices in each industry because of differing elasticities of derived demand for the factor in [133];
11. disguised unemployment in agriculture relative to manufacturing in [20], [78], [95];
12. differential factor taxation or subsidy in [76], [77], [91], [111], [132],
13. factory legislation, social regulation, or policy control for 'prestige-cum-humanitarian' or other normative reasons in [20], [22], [39];
14. movements in union/non-union differentials in the business cycle in [120], [125], [151]; and
15. limited mobility with differential product growth in [54], [78], [104].

In Bhagwati's terminology [19], the first eleven types are caused by endogenous forces while (12) and (13) are policy imposed; all thirteen are distortionary in a static context, while (14) and (15) involve both a distortion and comparative statics, raising the possibility of immiserizing growth Further, in dynamic systems, differentials may be observed during the adjustment process between two equilibrium points. This problem will be discussed later

Differentials, regardless of their cause, have two major effects the first is on the economic *structure* while the second is on *welfare* The structural effects are independent of whether the differential is distortionary. Bhagwati has suggested the following classification of the structural effects of a differential:

(i) the *output* effects,

(ii) the *shrinkage* of the production possibilities curve due to operation off the efficiency locus in capital-labour space,

(iii) *non-tangency*, i.e. non-equivalence of the marginal rates of transformation and substitution in product space,
(iv) *convexity*, i e. the possibility that the production possibilities curve becomes convex to the origin because of the differential,
(v) the *factor market* effects, including possible reversal of the product factor intensities, and
(vi) the *trade* effects

For convenience, these effects will be referred to in the text only by the word which is italicized The welfare and policy implications of a differential, on the other hand, depend on whether it is distortionary or not In most of the cases to be considered, we shall assume that the differential causes a distortion in order to pursue the welfare question While one-factor models are considered, most of the analysis pertains to two-factor models where we make the traditional assumptions of two factors in fixed total supply and immobile internationally, two goods, perfect competition in both product and factor markets, increasing opportunity costs between the two products, constant returns to scale in production, and no externalities The assumption of perfect factor markets where prices are flexible and the factors are perfectly mobile between industries will be relaxed in what follows

In Section II, factor price rigidity and factor immobility are considered Because the bulk of the recent literature deals with factor price differentials, the next three sections are devoted to them · Section III is an introduction and examination of the one-factor model, Section IV considers the structural effects in two-factor models, and Section V deals with the welfare effects Section VI is a brief review of empirical work and Section VII contains some concluding remarks

II. Factor price rigidity and immobility

A. Factor price rigidity

Haberler [52] and Johnson [77] have discussed the introduction of trade into an autarkic economy which is initially Pareto optimal In addition to trade, they added the condition that factor prices must remain rigid at their original level. This rigidity may be caused by a combination of institutional forces such as minimum wage legislation, governmental regulation or control, labour unions, or other forces. In the new equilibrium, production of the import-competing good must cease, given the assumption of constant returns to scale. Since all of the intensive factors in the import-competing sector cannot be absorbed by the expansion of the export industry, some of it will be unemployed in the new equilibrium The Stolper–Samuelson [146] analogue is that trade expansion has hurt

a portion of the import-competing industry's intensive factor via unemployment rather than through adverse price changes.

This result can be illustrated geometrically in Fig 1. Line L_0 (whose slope is P_x/P_y)[1] represents relative product prices in autarky. production and consumption occur at P_0 The introduction of trade increases P_x/P_y to L_2, moving production to P_2 and consumption to C_2 when no distortions

FIG. 1

are present However, with factor price rigidity, the new production point cannot be at P_2 but must occur between P_3 and C. The precise point depends on which good is the numeraire because of the index number problem. Haberler and Johnson conclude that free trade may actually reduce welfare in this situation. Bardhan [8], Batra and Pattanaik [10], and Kemp [85] have questioned the allocation of the net welfare effects of two exogenous changes (the introduction of trade and the introduction of the distortion via rigid prices) to one of the changes alone, i e trade We know that the introduction of trade alone cannot decrease welfare and

[1] Throughout this paper I shall ignore signs when referring to slopes of curves, demand elasticities, and elasticities of substitution Thus, an increase in the slope of L_0, for example, means a greater absolute slope rather than a greater algebraic slope.

the introduction of a distortion alone must reduce it. The effects of these two actions should be separated. Of course, no distortion is implied if factor prices are rigid because the supply of a factor is perfectly elastic at a given price (for the economy as a whole, this is precluded by our assumption of a fixed total supply of each factor).

Eckaus [39] and others have discussed the case where fixed coefficients in the production of both goods coupled with certain product price ratios imply a zero price for one of the factors of production, such as labour Societal constraints which require a non-zero wage impose a differential between the actual wage (which may be the same in all industries) and the shadow wage (which equals zero for the economy as a whole) Bhagwati [17] has used the Samuelson factor price equalization diagram to illustrate the unemployment effects of such a divergence between the actual and shadow wage in the more general case of a labour-surplus economy where the production functions permit smooth substitutibility To remedy this distortion, he recommends a tax-cum-subsidy on factor use in all sectors as the first-best policy; next in order of preference would be a tax-cum-subsidy policy on the domestic production of importables or a tariff on imports.

Policy remedies aimed at factor use are also first-best in Lewis's case [95] where wages are identical in all industries but they are above the marginal product of labour in one sector—agriculture Both shrinkage and non-tangency along the transformation curve contribute to the decline in welfare which accompanies this case.

Harris and Todaro [62] have considered theoretically a more recent problem plaguing less developed countries. i e. continued rural to urban migration in spite of positive marginal products in agriculture and substantial urban unemployment. They find that such behaviour is rational given urban minimum wage levels above agricultural earnings, government hiring and wage subsidies, and that migration responds to urban-rural differences in *expected* earnings Empirically, the correlation between regions of high unemployment and high wages has been observed by Hall [57] in the United States and Bryce [28] in Panama

B Factor immobility

We consider next the case where factor prices are flexible but the factors are immobile between industries. Cairnes [30] explored this possibility in a *one*-factor model with his famous theory of 'non-competing groups' of labour. He felt that all workers could be classified into four skill categories [30, pp 72–3] within which there was perfect substitutibility but between which there was none An analytical solution to this problem is to divide labour into four new factors and redefine the production

functions (see [31], [42], and [118]) Ohlin [119], however, felt that in the long run the different qualities of labour are not necessarily separable since high wages in skilled groups induce entry from lower-skilled categories

On the other hand, we can also assume that there are inter-industry barriers to labour mobility even though labour is of identical quality in both industries in a one-factor, two-good model By assuming that the economy is Pareto optimal in autarky and that factors will not move out of their autarky positions when trade is introduced, then with trade, we observe perfect factor immobility, and no change in the production of either good Furthermore, the Ricardian pattern of trade follows the country exports the good whose world price exceeds the autarky price The differential between wages in the two industries after trade, reflecting the failure of factors to move, causes the non-tangency between the new price line and the old transformation curve Welfare improves with trade in spite of the immobility. However, it would improve still further if labour became mobile and the system were permitted to specialize in the export product Viner [156, p. 124] gave an intuitive argument for the price and wage effects of trade in 1932 '[In this case], trade between the two countries will result only in changes in the relative prices of the two commodities, [and] in relative wages in the two occupations. '

The problem of factor immobility in *two*-factor models with international trade has been investigated by Haberler [52] and Johnson [77]. For simplicity, both assume that the country faces given terms of trade in world markets They then compare a country before and after trade to see if it is possible for the country to gain from trade when its domestic factors are completely immobile between industries In Fig 1, the country produces and consumes at the Pareto optimal point P_0 in autarky. If domestic factors were perfectly mobile, we find as before that the introduction of trade increases the price ratio P_x/P_y from L_0 to the international terms of trade denoted by line L_2, production moves to P_2 and consumption to C_2 in the Pareto optimal equilibrium. The gain from trade is illustrated by movement from community indifference curve U_0 to U_3

However, if factors are completely immobile, it is still possible to gain from trade If the increase in P_x/P_y from L_0 to L_1 is accompanied by flexible factor prices, then we remain at point P_0 in Fig. 1 and the gain from trade is from U_0 to U_2 Johnson [77, pp 13–14] notes that the total gain from trade (U_0 to U_3) can be broken into two parts the 'consumption or exchange' gain (U_0 to U_2) and the 'production or specialization' gain (U_2 to U_3). Thus, Haberler and Johnson showed, in effect, that the distortion introduced via factor immobility merely eliminates the production gain, the consumption gain remains when factor prices are flexible. The methodological criticism noted earlier ([8], [10], [85]),

applies here as well. The gain from trade is actually from U_0 to U_3; however, the introduction of the distortion via immobility results in a loss from U_3 back to U_2. Thus, the effect of both the exogenous changes is the movement from U_0 to U_2.

We note the following structural effects in this case. First, both Haberler and Johnson showed that the transformation curve in Fig. 1 shrinks to DP_0P_3, so that shrinkage affects all points on the transformation curve except the actual equilibrium point P_0.

Second, we know that the prices of both factors in the export industry (X) increase in proportion to the change in relative commodity prices when trade is introduced. This can be shown by letting P_x, P_y and w_x, w_y, r_x, r_y denote commodity and factor prices in X and Y and X_k, X_l and Y_k, Y_l denote the marginal productivities of capital and labour in X and Y. We assume that the traditional first order conditions hold in pretrade equilibrium before the rigidity is introduced

$$\frac{P_x}{P_y}\cdot\frac{X_l}{Y_l}=\frac{w_x}{w_y} \tag{1}$$

$$\frac{P_x}{P_y}\cdot\frac{X_k}{Y_k}=\frac{r_x}{r_y}. \tag{2}$$

With perfect competition, the right-hand side of both equations equals one, so that the system corresponds to P_0 in Fig 1 where L_0 is tangent to BP_0C. With factors immobile, the average physical productivities of the products can be treated as if they are constant. Thus, if P_x/P_y increases 20 per cent when trade is introduced, then w_x/w_y and r_x/r_y must increase by 20 per cent, so that in the new equilibrium $w_x = 1.2\,w_y$ and $r_x = 1.2\,r_y$.

This can be viewed in another way From the zero profit condition we can write

$$\frac{P_x}{P_y}=\left[\frac{Y}{X}\right]\cdot\left[\frac{w_xL_x+r_xK_x}{w_yL_y+r_yK_y}\right]. \tag{3}$$

By assumption, Y, X, L_x, K_x, L_y, and K_y all remain constant after trade is introduced so that a 20 per cent increase in P_x/P_y implies a 20 per cent increase in the ratio of X to Y's relative factor rewards (the second bracket on the right-hand side of (3)) We get the Cairnesian result that the return to these industry-specific factors are determined exclusively by changes in demand (the supply of the factors is perfectly 'inelastic') The Stolper–Samuelson [146] analogue would go something like this. increased relative export prices with immobile factors benefit all factors in the export industry proportionately, instead of just the intensive factor, as a result, a constant multiplicative differential is introduced in the price of X's factors relative to the price of identical factor services in Y.

Thus, the structural effects are formally similar to those in the one-factor case: an identical differential is paid for both factors in the export industry; the factor market equilibrium point, by definition, remains constant; there is no shrinkage of the transformation curve at the equilibrium point; the pattern of trade corresponds to the Heckscher-Ohlin theorem (if it would have without immobility); and the non-tangency issue is irrelevant since the marginal rate of transformation is indeterminate at the corner of the transformation curve DP_0P_3 at P_0 (although there is non-tangency between the new price line and the slope of the old transformation curve BP_0C) The case of factor immobility is important empirically it is relevant, for example, to the problem of the 'dual economy' discussed in much of the economic development literature.

Finally, following Johnson [77], we can combine factor *immobility* and factor price *rigidity* in four ways to yield the following results when trade is introduced (all are illustrated in Fig 1) We assume that trade causes P_x/P_y to increase

(i) If both factors are immobile and both factor prices rigid in terms of X, production of Y ceases and both of its factors are completely unemployed while production of X remains constant at point P_3

(ii) If both factors are immobile and the price of only one of them is rigid, then production of Y will not cease completely so that production will occur at some point such as G along the line P_0P_3

(iii) If one factor is mobile with a rigid price and the other is immobile but with a flexible price, the system will arrive at a point to the right of P_0P_3 and not above point G

(iv) If one factor is mobile with a flexible price and the other is immobile with a rigid price, the system will arrive at a point somewhere to the right of the line P_0P_3

These cases are also considered at some length in Batra and Pattanaik [10] In all of the cases where factor prices are rigid, the results depend on whether the rigid factor price is measured in terms of X, Y or some constant utility combination of the two

Jones [82], however, finds that the restricted entry form of factor market distortion does not result in non-tangency between the product price ratio and the transformation curve and the latter does not become convex while factor price differentials usually do generate non-tangency One difficulty with the differential model is that it may be difficult to maintain factor price differentials over long periods of time without restricted entry

Hemming and Corden [65] have examined optimal policy alternatives

in cases where output changes result in unemployment, the latter being **necessary for labour movement This work has been extended by Ramaswami [127] and Corden [35] In another context, Ramaswami [126] examines *policy-induced international factor immobility*** *In considering optimal restrictions on factor movements, he finds that optimal taxation* of the immigration of the *scarce factor is superior to optimal restriction* of the emigration of the abundant factor This was extended by Webb [159] to the case where two governments negotiate such taxation in a bilateral monopoly situation

III. An introduction to factor price differentials

A Types of differentials

The most widely discussed cause of factor market distortions is the case where factors are relatively mobile and their prices flexible and yet a differential exists in the price of a factor in different industries As mentioned earlier, this case is not always clearly distinguishable from the case where immobility exists For example, labour unions frequently exploit their monopoly power by a combination of price setting and restricted entry, the latter being a special form of immobility

Differentials may be observed in static or comparative static systems and in systems in the dynamic adjustment process between two equilibria We noted early in the paper a number of causes of non-dynamic differentials We will discuss briefly the *dynamic* variety They are harder to classify as distortions since the size of the differential depends on the time period chosen, the speed of convergence (assuming stability), the size of the initial shock, and how recent the system was jolted by some parametric shift The dynamic differential is another case of partial factor mobility where adjustment to change is not instantaneous A non-explosive cobweb model with lagged supply and demand reactions generates an observable differential if the period of observation is small or no differential if it is sufficiently long Hagen's [54] use of more rapid demand expansion for manufactures than for agriculture to generate a factor distortion has been criticized by Kenen [88], Koo [91], Fishlow and David [42], and Bhagwati and Ramaswami [20] as 'illegitimately superimposing a dynamic argument upon a comparative statics framework' [20, p 48] Johansen [75] has a dynamic model which explains interindustry wage differentials in terms of different productivity growth

Recent papers, following Inada [74] and Uekawa [152], have explored growth models in which one factor (Herberg [68]) or all factors (Herberg and Kemp [70]) are imperfectly mobile This allows generation of endogenous differentials, which disappear in the steady state The empirica

question of growth with regional income differentials has been analysed by Fukuchi and Nobukuni [48] for Japan and others

In less formal models, one can imagine both 'demand-pull differentials' and 'cost-push differentials', generated by sectoral shifts in either the demand for or supply of particular factors (with the other curve being less than perfectly elastic) Regardless of the initial cause, the disparity in prices will not remain as a permanent differential unless some economic or non-economic force acts to maintain it Schlesinger [137] for example, has attributed union success to its alteration of the labour supply function, consequently, 'excessive' wage demands would classify unionization as a cost-push differential But the union differential may not be maintained indefinitely without restricted entry since free entry would erode the differential On the other hand, discrimination by race, sex, or age might be introduced through a demand-pull differential in favour of certain segments of the labour force, which is maintained on largely non-economic grounds These assumptions of restricted entry or non-economic forces are not required to justify differentials caused by factor taxation which differs by sector, since after tax returns are everywhere equal

In short, one should examine very carefully the way in which a differential is introduced, whether it is narrowing through the traditional dynamic adjustment process and if not, what assumptions are required to maintain it More systematic analyses should be done on the *policy* implications of the time dimension, such as the present discounted costs of dynamic endogenous differentials and the stability properties of tax-cum-subsidy policies designed to close endogenous differentials more rapidly

We turn now from differentials caused by dynamic forces to static models, in which differentials are exogenous to the system In the next section we shall examine static differentials in one-factor models

B. *Differentials in one-factor models*

The earliest works on factor price differentials dealt primarily with the Ricardian case of only one factor labour See for example the discussions in Cairnes [30], Taussig [149] Ohlin [118] Viner [156] and Hagen [54] Hagen [54, pp 504 5] has the best discussion of the one-factor two-good model We have reproduced one of his diagrams in Fig 2 The transformation curve BC is linear because of the one-factor constant returns assumption With no distortion, the economy will produce and consume at A in autarky If we introduce a differential such that wages in X are double those in Y, the new domestic price ratio P_x/P_y will equal the slope of BC', which is twice as steep as BC This can be formalized as follows Using

the fixed coefficient production functions $X = aL_x$ and $Y = bL_y$, profit maximization yields the non-specialized first order conditions $P_x = w_x/a$ and $P_y = w_y/b$ where w_x and w_y are the wages in X and Y. Thus,

$$\frac{P_x}{P_y} = \frac{w_x}{w_y} \cdot \frac{b}{a}.$$

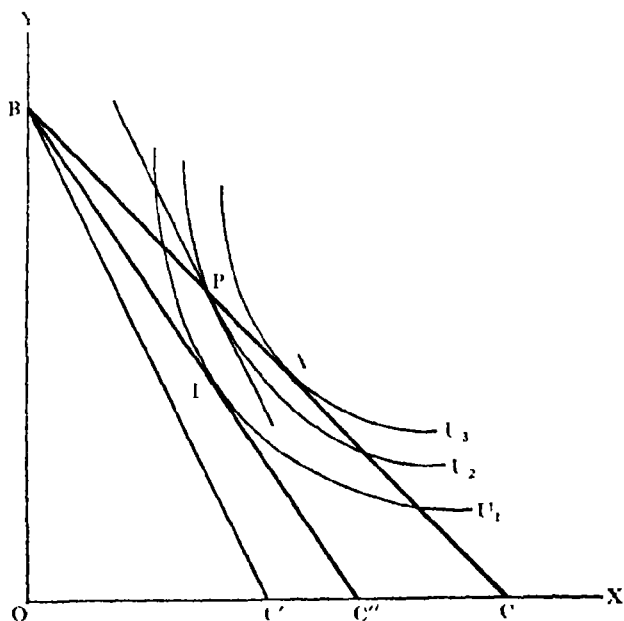



Fig. 2

Without distortion, $w_x/w_y = 1$, but when we double (w_x/w_y), then P_x/P_y doubles. Thus, the original price ratio along BC in Fig 2 is $[b/a]$ while the slope of a price line such as BC'' is $2[b/a]$.

In the absence of trade, the economy will produce and consume at P when the distortion is present. At that point, the marginal rate of substitution, which equals the price ratio corresponding to the slope of BC'', does not equal the marginal rate of transformation (the slope of BC'). If the country initiates trade and faces fixed international terms of trade, slope BC'', which lies between the slopes of BC and BC', the country will specialize in Y, export Y, and consume at T, which in this case is inferior to both P and A. Hagen was using this analysis to show that protectionism can be superior to free trade, i.e. that the utility level at P with a prohibitive tariff and factor distortion is superior to the free trade level at T.

Notice also that the pattern of trade depends on the size and direction of the differential. If the distorted domestic price ratio had been less than the international terms of trade (slope BC''), then the country would have

specialized in X rather than Y. Using the formula developed above, we can say, in general, that if $[P_x/P_y]_I$ denotes the international terms of trade

$$\left[\frac{P_x}{P_y}\right]_I \gtrless \frac{w_x}{w_y} \cdot \frac{b}{a} \text{ implies specialization in } \begin{Bmatrix} X \\ Y \end{Bmatrix}.$$

Clearly, w_x/w_y can be manipulated by the differential to cause specialization either way In summary, the differential reduces output of the industry paying the differential, causes non-tangency in autarky, and can reverse the pattern of trade.

Models with one factor and more than two products were discussed by Cairnes [**30**] in his controversy with the classicists Cairnes argued that if labour is divided into non-competing groups with each group aligned with a specific product, wage differentials will be determined exclusively by the demand for each product Marshall and Taussig [**149**, pp 53-4] criticized Cairnes for ignoring the long-run supply response of labour (reproduction would increase within each non-competing group whenever its wages rose above the subsistence level) Taussig [**149** p 57] went so far as to argue that international trade 'is not likely to modify the alignment of grades [of labour] within a country', although that alignment could certainly affect international trade, as in his cases of pre-World War I German chemical exports and United States iron and steel exports Taussig showed that the classical explanation of the pattern of trade still holds if the structure of relative wage distortions is the same between countries

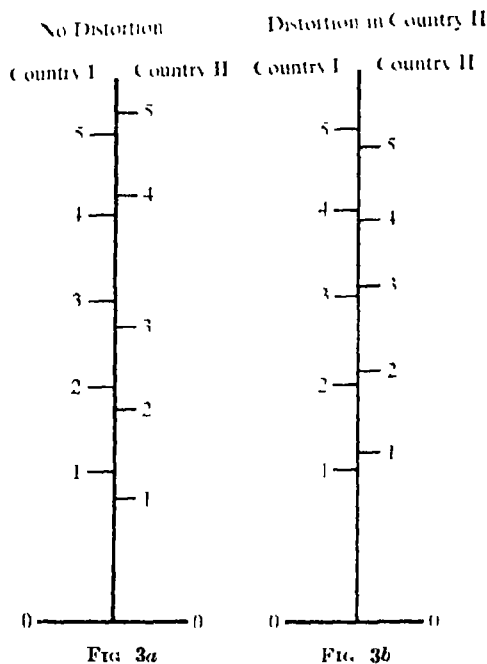
Samuelson [**135**] and others have discussed comparative advantage in a one-factor, many-good model without wage distortion an explanation of the pattern of trade in this one-factor model *with* wage distortion can be developed by borrowing the Mangoldt-Edgeworth technique discussed by Viner [**158**, pp 458-62] We shall consider only two countries The technique is general enough to consider cases where both the level and structure of wages differ by country but we are interested here only in the structural aspects (this is oversimplifying since the levels of wages are determined partly by the reciprocal demand of the two countries for each other's products) We noted earlier that the cost of good i in a one-factor model (where a_i represents the constant output labour ratio and w_i represents that wage paid by industry i) can be written

$$P_i = w_i/a_i \qquad i = 1, \ldots 5 \tag{4}$$

Assume initially that there are no factor price distortions so that w_i is the same for all five industries We can plot the logarithms of $P_1 \ldots P_5$ for country I from a fixed point 0 in Fig. 3*a* on the left side of the line Similarly, the logarithms of the prices, $P_1, \ldots, P_5$ for country II, measured from the same point 0', are plotted on the right side of the line in Fig 3*a* If the wage levels in country II are greater (less) than those in I, 0' will lie above (below) 0

We have ruled this out by assumption. We can now determine which products will be exported by each country. Since the costs of goods 1, 2 and 3 are lower and 4, 5, higher in country II in Fig. 3*a*, then country II exports the former and country I the latter.

The Mangoldt–Edgeworth diagram can be used to show that when wage differentials are introduced, it is possible to obtain almost any pattern of trade. For example, if the relative wages of goods 1, 2, and 3 are increased

FIG. 3*a* FIG. 3*b*

and those of 4 and 5 fall in country II with the introduction of the differentials, then the pattern of trade can be reversed as in Fig. 3*b*. The same analysis can be performed using equation (4). This approach also illustrates Taussig's point that if the wage distortions are the same between countries the pattern of trade is unaffected.

IV. Differentials in two-factor models: the structural effects

A. Preliminary remarks, shrinkage, and non-tangency

Two important theoretical articles in the past decade on the structural effects of factor price differentials were published by Harberger [**60**] and Johnson [**78**]. Harberger's paper was concerned exclusively with the incidence of the U.S. corporate income-tax, and was followed by Mieszkowski [**111**]. After some delay, Johnson's paper stimulated an avalanche of research on the mathematical properties of general equilibrium models in which factor price differentials are present. At least six theoretical

pers were written independently which explored factor price differentials d their effects on various properties of an economic system Bhagwati d Srinivasan [23], Herberg and Kemp [69], Johnson and Mieszkowski 0], Lloyd [97], Magee [100], and Mundlak [115] While each paper was ncerned with slightly different problems, there was virtual unanimity results where identical problems were considered These papers cited others by Batra and Pattanaik [11], Herberg, Kemp, and Magee 1], and Jones [82], among others In this subsection, we shall explore e shrinkage and non-tangency issues

First, we shall examine the problem of shrinkage of the production ssibility curve and its non-tangency with the product price line We ted earlier that factor-price differentials in two-factor models may place e system below its optimum transformation curve in product space (the *rinkage* effect) if the factor market equilibrium is not on the efficiency cus in the Edgeworth–Bowley box [1] Point A in Figs 4a and 4b indicates distortion while point D may indicate, for example, that X is paying a latively higher price for capital than Y Point D is on a transformation rve BDC in Fig 4b which is inferior to the optimum curve BAC If output Y is held constant at Y_m in Fig 4a (equal to OM in Fig 4b) removal of the stortion (allowing product prices to change appropriately) implies movement from point D to E (in Fig 4a), which increases output of X from X_d X_e (or from MD to ME in Fig 4b). Extending this argument to every point ong the outer transformation curve we can show that for any degree of stortion, the distortion curve BDC lies everywhere inside BAC, except the two specialization points B and C Eckaus [39, p 373] has called e outer curve the 'technical transformation curve' and those inside of 'market transformation curves' since they are generated by factor arket distortions We shall follow Fishlow and David [42] and others lling in them simply 'outer' and 'inner' transformation curves

Factor price differentials can break the equivalence of the marginal chnical rates of substitution between factors and the marginal rates of bstitution and transformation between products Next, we shall examine en this does and does not occur Again, we consider the two-good, o-factor model discussed in equations (1) and (2) We can rewrite the st-order conditions for profit maximization as follows

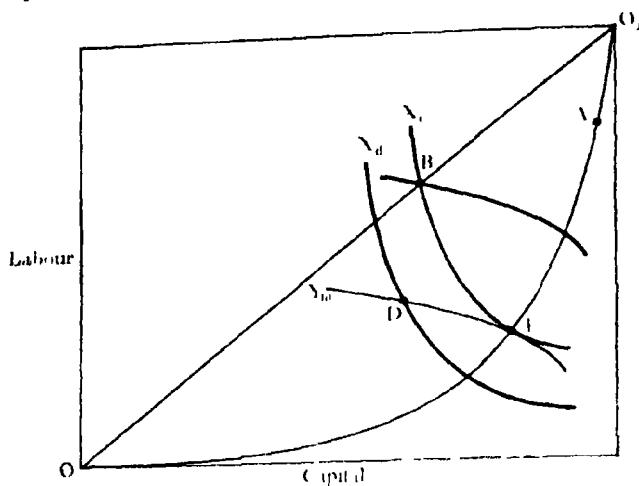
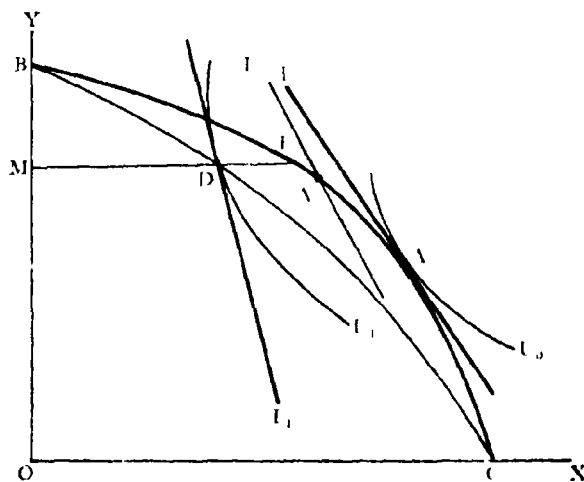
$$P_x X_l = w_x \tag{5}$$

$$P_y Y_l = w_y \tag{6}$$

$$P_x X_k = r_x \tag{7}$$

$$P_y Y_k = r_y . \tag{8}$$

[1] Jorgo Marquez has shown that the shrinkage phenomenon is a second order effect ting from no distortion

A distortion occurs whenever a differential exists such that $w_x \neq w_y$ or $r_x \neq r_y$. Following Hagen [54, pp 507–8], we can take the total derivati[ves] of the production functions for both goods, replace the marginal produ[c-]

tivities with their equilibrium values from equations (5) to (8), use th[e] endowment constraints and write

$$-\frac{\mathrm{d}Y}{\mathrm{d}X}=\frac{P_x}{P_y}\cdot\left[\frac{r_y\mathrm{d}K_x+w_y\mathrm{d}L_x}{r_x\mathrm{d}K_x+w_x\mathrm{d}L_x}\right]. \qquad (9)$$

The marginal rate of transformation (MRT) and the marginal rate [of] substitution (MRS) differ if the bracketed term in (9), which we shall [call] 'the non-tangency factor', differs from one

If the cost of both factors is identical in each industry, equations (5) to (8) yield

$$\frac{X_l}{Y_l} = \frac{X_k}{Y_k} \tag{10}$$

and the bracketed term in (9) equals one so that the system is Pareto optimal and no distortion exists. Thus, we get a factor market *distortion* and *shrinkage* if the equality in (10), is broken, and *non-tangency* if the non-tangency factor in (9) differs from one. We shall examine three possibilities

first, non-tangency but no shrinkage,
second, shrinkage but no non-tangency; and
third, both shrinkage and non-tangency

First, consider non-tangency but no shrinkage. If industry (X) faces an identical differential t (> 1) for both factors, there is no distortion in the factor markets since from equations (5) to (8),

$$\frac{X_k}{X_l} = \frac{tY_k}{tY_l} \tag{11}$$

and the differentials cancel Since we remain on the efficiency locus in the Edgeworth–Bowley box, we also remain on the outer transformation curve BAC in X–Y space The product markets are distorted, however, since from equation (9),

$$MRT = \frac{1}{t} MRS. \tag{12}$$

Thus, we have violated the marginal equivalences in product space so that price line L_0' will intersect BAC at a point such as A' from above in Fig 4*b*, with the non-tangency factor less than one This case is similar to that of pure monopoly

Second, we consider shrinkage but no non-tangency If industry X pays a higher price for $K2(t = r_x/r_y > 1)$ but a lower price for $L2(s = w_x/w_y < 1)$, then the factor market is distorted since

$$\frac{X_k}{X_1} = \frac{t}{s} \cdot \frac{Y_k}{Y_1} \tag{13}$$

with $(t/s) > 1$ The empirical possibility of this case in less developed countries has been considered by Myint [**117**], among others The product market equivalences may or may not be distorted, however, since the $MRT \lessgtr MRS$ depending on whether the net effect of s and t causes the non-tangency factor in (9) to be $\lessgtr 1$ In the special case where the non-tangency factor equals one, we have an example in which the marginal factor equivalences are broken but the marginal product equivalences are not. Thus, the price line would be tangent to the inner transformation curve at the production equilibrium point.

Third, we consider both shrinkage and non-tangency. This includes situations in the second case, just considered, in which the non-tangen factor does not equal one. It also includes situations in which only one the two-factor markets is distorted If industry X has to pay a high price for one of its factors, such as capital ($r_x > r_y$), the system mov above the efficiency locus to a point such as D in Fig 4*a* so that

$$\frac{X_k}{X_1} = t\left[\frac{Y_k}{Y_1}\right] \quad (1$$

where $t = (r_x/r_y) > 1$ The marginal technical rates of factor substitutio now differ by t Similarly, from (9) we know that the non-tangency fact is less than one so that $MRS > MRT$, i e the relative market price of . is greater than its opportunity cost because of the higher factor mark costs to X induced by the differential The slope of the price line $L_1 = MR$ is greater than MRT, the slope of the inner transformation curve BDC' a D in Fig 4*b* Thus, when there is only one differential paid by on industry, the marginal equivalences are broken in both the factor an product markets Since much of the existing literature deals with th case of the single differential, we shall now explore it in some depth

B The factor market and supply response

We start the discussion with the factor market, since important result for the entire system depend upon certain features of the factor mark Assume that products X and Y are produced by linearly homogeneou production functions and that the capital–labour ratios in each industr are denoted by k_x and k_y We assume in the absence of distortion tha product X is capital intensive The distortion is introduced by assumin that the capital market incorporates a factor price differential t, defin by

$$r_x = t\, r_y \quad (15$$

while the labour market does not ($w_x = w_y$) Following Johnson [78], we know from equation (14) that if $t > 1$ (< 1), we are above (below) th non-distorted efficiency locus $O_x EO_y$ in Fig 4*a*, since the left-hand side o equation (14) is the slope of the X isoquant and the bracketed term on th right-hand side is the slope of the Y isoquant Any locus of points alon which t is constant and unequal to 1 is called a 'distorted efficiency locus (equilibrium product prices, P_x/P_y, vary along this line) Correspondin to every distorted efficiency locus, there is an 'inner' or shrunken pro duction possibility curve in the output space Any sequence of equilibrium points along which relative product prices are fixed while t varies is calle a 'distortion equilibrium locus'.

Once we leave the special case where $t = 1$, we must distinguish tw types of factor intensities those in the physical sense and those in th

value sense (see Jones [82]). Let PH represent the 'physical' definition of relative factor intensity, i.e the difference in the capital–labour ratios in the two industries.

$$PH = k_x - k_y. \tag{16}$$

Similarly, let VA represent the 'value' definition of relative factor intensity, i.e. the difference in the shares of capital relative to labour in the two industries.

$$VA = \frac{r_x K_x}{w_x L_x} - \frac{r_y K_y}{w_y L_y} = (r/w)_x k_x - (r/w)_y k_y \tag{17}$$

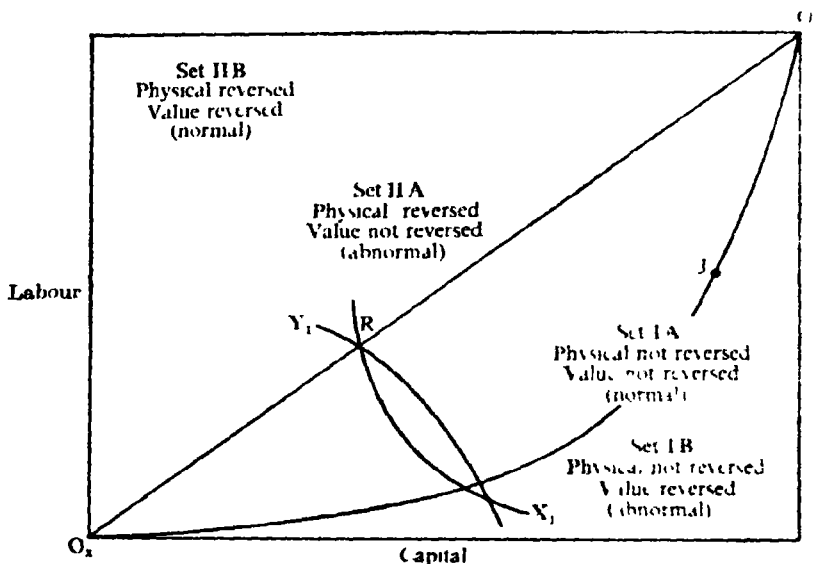



Fig. 5

In the absence of distortion (when $t = 1$), both PH and VA always have the same sign, so that along the non-distorted efficiency locus in the Edgeworth–Bowley box, the physical factor intensities and the value factor intensities of the products always correspond

Johnson [78] has noted that if an industry pays a differential for its physically intensive factor, the *physical* factor intensities of the products can reverse If production isoquants are homothetic, the differential t must exceed the ratio of the slopes of the X to the Y isoquants along the diagonal in the Edgeworth–Bowley box (e g. the relative slopes at point R in Fig 5) We shall denote by sets I and II non-specialized factor market equilibria for which the physical factor intensities are not reversed or are reversed, respectively In Fig 5, set I corresponds to non-specialized points below the diagonal and set II to non-specialized points above the diagonal Magee [101] showed that it is impossible to move from set I to set II by merely changing the differential Herberg, Kemp, and Magee

[71] extended this, proving that the set of relative product prices giving non-specialized equilibrium below the diagonal in the Edgeworth–Bowley box and the set of relative product prices giving equilibrium above the diagonal are mutually exclusive Geometrically it is clear that the sets of t's corresponding to non-reversal in set I in Fig 5 does not overlap with the t's in set II ([71], [78], [80], [101], and [115]) All of the values of t and P_x/P_y in Set II exceed the values of t and P_x/P_y in set I in Fig. 5 [71].[1] In order to cross the diagonal and remain non-specialized, *both* the differential and relative product prices must change

We turn now to the effect of differentials on the relative *value* factor intensities Bhagwati and Srinivasan [23], Herberg and Kemp [69], Jones [82], Lloyd [97], Magee [101], and Mundlak [115] have all shown that factor price differentials can reverse relative value factor intensities We shall denote by sets A and B non-specialized factor market equilibria for which the original value factor intensities are not reversed or are reversed, respectively. For many production functions, it is possible for the value factor intensities to reverse on both sides of the diagonal, so that we get the configuration shown in Fig 5 One comment which relates Fig 5 to factor markets in general is that all non-specialized points between the non-distorted efficiency locus $O_x JO_y$ and the diagonal $O_x RO_y$, both inclusive, are members of set A [101] Otherwise, the regions denoting sets A and B are purely stylized—they are shown only to denote logical possibilities and not a particular concatenation of factor market equilibria.

The importance of the physical value distinction for the entire system (not just factor markets) is explored by Jones [82] His argument runs as follows Physical factor proportions link changes in the real variables of the model, the community's outputs, and factor endowments The value ranking given by distributive shares serves to link the financial variables, the prices of outputs, and rental for the service of inputs The relationship between changes in physical variables (e g outputs) and financial variables (e g output prices) depend crucially on whether physical and value rankings of factor proportions correspond The uses of the words 'financial' and 'real' do not imply the existence of any monetary assets in the model, but rather the distinction between relative price and share variables on the one hand and quantity variables on the other The two definitions of factor intensity correspond (do not correspond) if PH and VA in equations (16) and (17) have the same (opposite) sign

For example, many of the recent papers ([11], [23], [69], [71], [82], [97] [101], and [115]) find that an increase in P P_x/P_y (with the differential held constant) increases the supply of X if the physical and value in-

[1] Intuitively, as t increases the X industry is penalized as it moves toward Set II, to remain non-specialized, P_x/P_y must be raised to offset the higher

nsities correspond and reduces X if they do not. Consequently, the upply response of outputs to changes in P is *normal* if, and only if, neither he physical nor the value intensities are reversed (set IA in Fig. 5), or both the physical and value intensities are reversed (set IIB) and *abnormal* f, and only if, relative factor intensities have reversed in the physical ense only (set IIA) or relative factor intensities have reversed in the value ense only (set IB). Thus, the supply of a product is positively related to elative product prices in the normal cases; supply is perfectly elastic with espect to prices whenever the physical or the value factor intensities of he products are identical,[1] and supply is negatively related to prices in the abnormal cases.

The response of outputs to changes in the differential (with product rices constant) is the same (see [11], [71], [82], [101], and [115]). An ncrease in a differential paid by an industry on either factor reduces output (the normal case) if the physical and value factor intensities correspond nd increases output if they do not (the abnormal case).

Now that the output effects have been mentioned, we return to a more etailed analysis of factor market behaviour. Until very recently, writers in his area were either not aware of or failed to pursue the perverse possibilities osed by *non-correspondence of the physical and value factor intensities* Harberger [60], Johnson [78], Johnson and Mieszkowski [80], Mieszkowski [111]). Consequently, their contributions cover half of the logical possibilities — those in which the response of the system to changes in either elative product prices on the differential is normal.

Since all of the points along the non-distorted efficiency locus O_xAO_y n Fig. 6 are in the *normal* set IA, let us consider the effects of the introduction of a capital differential in industry X starting from point A such hat t increases from $t = 1$ to $t > 1$ (see [80] and [101]). First, if we hold utput of X constant, there is a 'substitution' effect away from capital in ndustry X, indicated by movement from A to E. The substitution effect auses the capital–labour ratio to decline in X and rise in Y. The curve O_xEO_y is a distorted efficiency locus, along which the ratio of the X to the isoquants equals the new value of the differential t. The system cannot emain at E, however, since negative profits exist in the X industry; thus, here is an 'output' effect, which causes a reduction in the production of X. Consequently, the new factor market equilibrium must be below E on he distorted efficiency locus O_xEO_y. The output effect causes the capital-abour ratios to increase in both X and Y (starting from point E). In erms of the effect on the capital–labour ratio in the industry paying the ifferential, the substitution and output effects work in opposite directions

[1] The reason is that the derivatives of outputs with respect to relative prices have zeros n their denominators when either PH or VA equal zero

when the differential is increased on the industry's intensive factor and in the same direction when it is increased on the non-intensive factor.

An important question is whether the new factor market equilibrium is above or below point G on the distorted efficiency locus Jones [**82**] and Magee [**100**] and [**101**] have shown that if relative product prices are unchanged from their value at point A (e g if the country does not possess

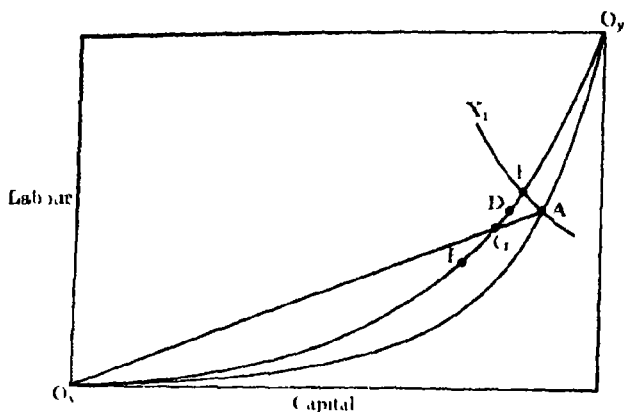



Fig. 6

monopoly power in trade), the new equilibrium with the increase in t must always be below point G, at a point such as F Thus, whenever an industry must pay a differential on its intensive factor, such as capital, with product prices fixed, *the output effect always dominates the substitution effect.* Capital labour ratios rise in both industries and by the homogeneity postulate capital is *worse* off in both industries The implications for labour unions or any factor with market power are clear if product prices are relatively fixed they should introduce differentials in sectors of the economy in which they are the non-intensive factor [**101**] A related result is Johnson's [**79** examination of minimum wage laws He finds that if they are applied to only one sector, labour may be better off in all sectors On the other hand a minimum wage law applied to all sectors will result in unemployment

If product prices are not fixed we get a third effect, called the 'price effect The increase in the price of the product paying the increased differential will mitigate part of the adverse output effect so that the new equilibrium will be above F on the distorted efficiency locus If the new equilibrium is below G, the previous factor welfare results still hold, if it is above G, then capital is better off in industry X and worse off in Y. The actual location of the new point depends upon demand (see Harberger [**60** Johnson and Mieszkowski [**80**], and Mieszkowski [**111**] for discussion of this point).

Returning to the situation in which output prices are fixed, and considering all cases, we can say that an increase in a differential causes a decrease (increase) in the relative price of the factor if it is the physically intensive (non-intensive) factor in the industry paying the increased differential when the physical and the value factor intensities correspond [1] This leads to an increase (reduction) in the physical proportions of that

FIG. 7

factor used in both industries The same statement holds for a decrease in the relative price of a commodity with any fixed differential This implies that if we continue to define factor intensities in a physical sense then the Stolper–Samuelson theorem holds in the normal case, but is reversed in the abnormal case ([11] and [101]) If, however, factor intensities are defined in the value sense, the Stolper Samuelson theorem holds in all cases [71] This follows Jones's [82] dictum that the value ranking given by distributive shares serves as a link between all financial variables (product prices and factor prices)

A useful theorem relating the factor market to the product market is that if both capital–labour ratios move in the same direction then output responses are unambiguously determined, since k_x and k_y move in the same direction, the knowledge of dk_x/dt or dk_x/dP is sufficient to determine dX/dt, dY/dt, dX/dP, and dY/dP, and the latter need not be calculated ([100] and [101]) This is illustrated in Fig 7 Starting from point L (where neither type of reversal has occurred), an increase in the differential

[1] The opposite holds when the factor intensities do not correspond

which X must pay for capital (regardless of whether $t > 1$ or $t < 1$ initia increases k_x and k_y so that the new equilibrium must be in shaded area Clearly, output of X falls and Y increases. *If the physical but not the va factor intensities of the products are reversed so that we start from point an increase in the capital differential paid by X results in a perverse resu namely, an increase in X and decrease in Y as the factor market equilibriu moves from N into area D.*

*With regard to factor supplies, we know that an exogenous increase in a country's endowment of a factor with distortion leads to a relative increase in the production of the good using that factor intensely in the physical sense Consequently, so long as we define factor intensities in the physical sense, Rybczynski's theorem holds without modification, regardless of the degree of factor market distortion and regardless of whether the distortion has reversed the value factor intensities of the product ([**82**] and [**101**])*

C. Physical and value factor intensity reversals

It is of some interest to know what characteristics of the production functions are associated with physical and value factor intensity reversals. The value of t required to cross the diagonal and give physical reversal diminishes as the product elasticities of factor substitution increase in size and as the non-distorted factor intensities of the products become more similar It is more difficult to generalize about value factor intensity reversals [1] They are impossible with Cobb–Douglas production functions since the relative shares going to capital and labour are invariant to changes in either prices or the differential (Herberg and Kemp [**69**])

If the physical factor intensities have not been reversed, value factor intensity reversals are possible only if an industry pays a differential on its non-intensive factor, which places the factor market equilibrium in the area on the side of the non-distorted efficiency locus opposite the diagonal in the Edgeworth–Bowley box ([**82**], [**101**], and [**115**]) Thus, value reversals are impossible if $t > 1$ in set I. In general, if t is held constant and P_x/P_y varies, the value factor intensities never reverse if S_x and S_y are constant and equal, reverse once if the elasticities of factor substitution, S_x and S_y, are constant and different, and may reverse more than once in all other cases [**71**] It is also true that for any given value of k_y (along which $(r/w)_y$ is constant by the homogeneity assumption), the value intensities never reverse if $S_x = 1$, reverse once if S_x is constant but unequal to one and may reverse more than once in all other cases [**71**]

Geometrically, the Samuelson–Johnson factor price equalization diagram can be used to extend these results (see [**100**] and [**101**] for the

[1] i e as the isoquants get flatter and their tangencies get closer to the diagonal

application of this diagram to factor price differentials) Specifica even in the case where an industry pays a differential on its non-intens factor in set I ($t < 1$ or below the non-distorted efficiency locus $O_x JO_y$ the Edgeworth–Bowley box in Fig 5), it is impossible for the value fac intensities to reverse if $S_x \leq 1$ and $S_y \geq 1$ The argument proceeds follows. Since the country does not possess monopoly power in trade,

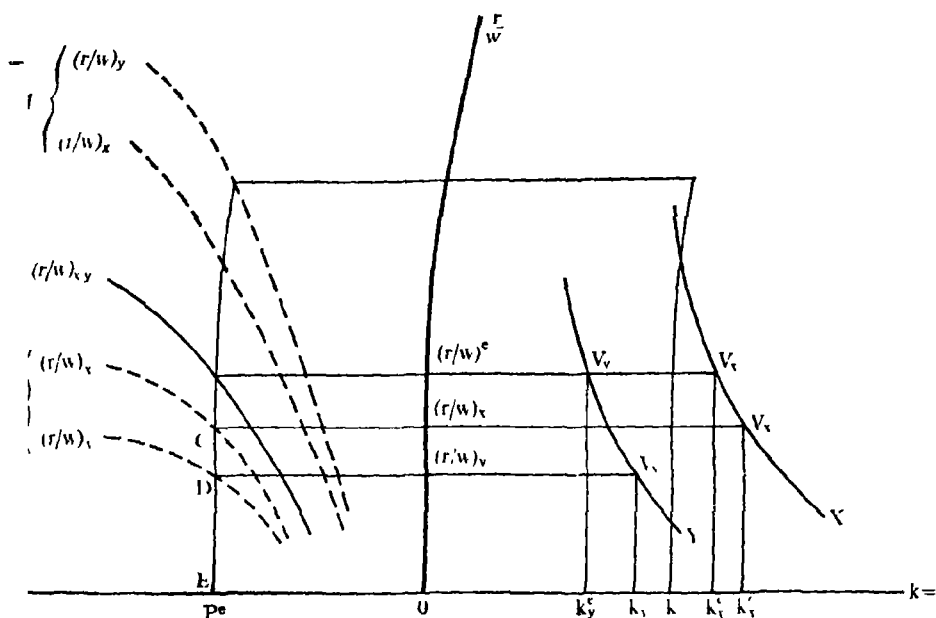



FIG 8

faces a fixed world equilibrium product price ratio, P^e The areas under the X and Y curves in the right-hand side of Fig 8 equal the ratio of the capital to labour factor shares in each industry, rK/wL The elasticities of factor substitution in the two industries, i.e the percentage change in K/L for a given percentage change in r/w, are equal to the elasticities of the X and Y curves

Observe that in the absence of distortion ($t = 1$), the physical and the value factor intensities of the two products must coincide. From equation (16) the difference in physical factor intensities at $(r/w)^e$ is

$$PH \equiv 0k^e_x - 0k^e_y > 0$$

while the difference in the value factor intensities from equation (17) is the difference in the areas under the X and Y curves, or

$$VA = \text{area } 0V_x - \text{area } 0V_y \equiv \text{area } k^e_y V_x > 0.$$

If t increases to some value greater than one, the unique link between

product prices and factor prices in the left-hand side of the diagram broken we now have a separate $(r/w)_i$ curve for each industry i. We sa earlier that at fixed commodity prices, the relative return to capital mu fall in both industries Thus, the two new curves lie below the old $t =$ curve and the ratio of the height of the X to Y curve must equal the ne differential ($t = CE/DE$). When the value and the physical factor ir

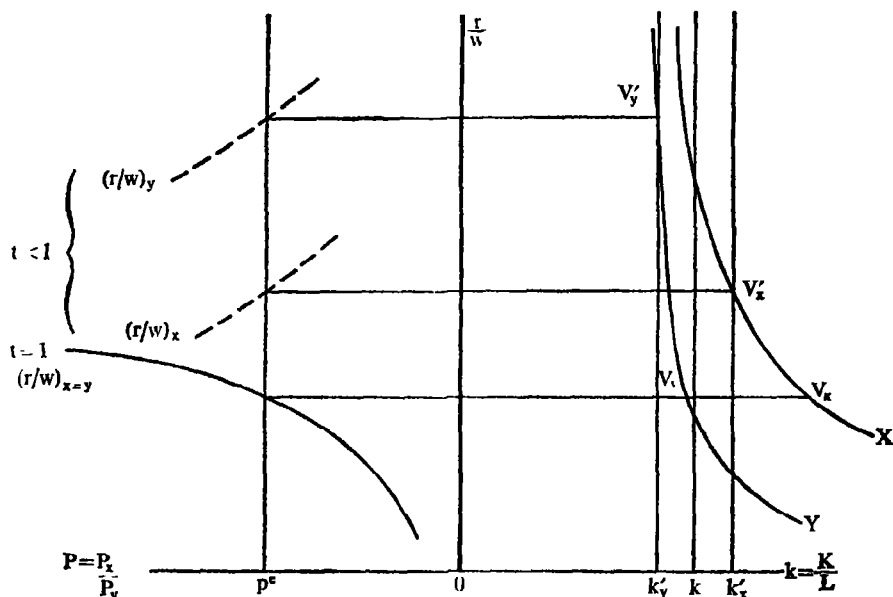



FIG 9

tensities correspond (do not correspond), the curves are positively (negatively) sloped. Whether capital increases or decreases its share of the value of output in each industry when t changes depends on the elasticities of factor substitution, S_i ($i = X, Y$) If $S_i < 1$, the share of capital in industry i falls as the area under the curve declines If $S_i > 1$, the share of capital increases

If industry Y, where capital is the non-intensive factor, is forced to pay a differential for capital ($t < 1$), these results are just reversed capital is better (worse) off in both industries and increases (decreases) its share of industry i's output if $S_i < 1 (> 1)$ If t decreases sufficiently, we get the case shown in Fig 9 where the value but not the physical factor intensities of the products have been reversed, since

$$VA_{t<1} \equiv \text{area } 0V_x' - \text{area } 0V_y' < 0$$

The dashed lines $(r/w)_i$ for $t < 1$ must be negatively sloped since the relationship $(r/w)_i = f_i(P_x/P_y)$ has the property $f_i' > 0$ without reversal and $f_i' < 0$ with reversal, as mentioned earlier.

In general, starting from any point in set 1A, reducing t moves points V_x and V_y up the X and Y curves in the right-hand side of Fig 8 [**101**]. Geometrically, it is clear that VA will not change signs as t is decreased if the area under the X curve is constant or increases while the area under the Y curve is constant or decreases, i e $S_x \leq 1$ and $S_y \geq 1$, which is what we set out to establish on page 24 Notice that the elasticities of substitution in this case do not need to be constant.

D The shape of the transformation curve

We have noted the possibility of perverse supply response to relative product price changes A related area which has drawn attention is the question of the shape of the transformation curve with differentials The problem has been examined by [**20**], [**23**], [**42**], [**69**], [**78**], [**82**], and [**97**], to name a few While this is of enormous theoretical importance, it is of little empirical importance In Mundlak's [**115**, p 531] words, 'The examination of the convexity of the transformation curve is of little interest What matters is the relationships between [outputs] and [prices] ' Ironically, this statement is true because of the extensive examinations of Bhagwati and Srinivasan [**23**], Herberg and Kemp [**69**], and Lloyd [**97**] into the convexity question Independently, they discovered the important result that normal and perverse output responses to price changes are not necessarily related to the concavity or convexity (from below) of the transformation curve Thus, normal output response is possible along convex portions and perverse response along concave portions of the transformation curve Jones [**82**] has extended the discussion, showing how the 'wedge' (i e the gap or the non-tangency factor between the price line and the distorted transformation curve) varies along the curve

Many propositions have been derived linking the shape of the transformation curve to the factor market and production functions We shall mention three

1 If a differential paid to the *intensive* factor (so measured in the absence of distortion) in an industry is large enough, it can cause the physical factor intensities to reverse Before reversal, the transformation curve lies outside a chord joining its end points, at the reversal value, the transformation curve coincides with the chord after reversal, the transformation curve lies inside the chord (Johnson [**78**])

2 In the Cobb-Douglas case, a differential paid by an industry on its *intensive* factor (above $O_x JO_y$ in Fig. 5) will be accompanied by a concave transformation curve with normal output response to price changes before physical reversal, and a convex transformation curve

with perverse output response after physical reversal (Johnson and Herberg and Kemp [69]).

3 If, in the Cobb–Douglas case, a differential is paid by an industry to its *non-intensive* factor (below $O_x JO_y$ in Fig 5), then all output relationships are normal, regardless of the concavity or convexity of the transformation curve, in fact, the transformation curve may be uniformly convex or uniformly concave or partly concave and partly convex, but it cannot change its curvature more than twice (Herberg and Kemp [69])

E International trade and factor movements

The Samuelson–Johnson diagram in Fig 8 illustrates the impossibility of international factor price equalization with differentials If the home factor market has a differential, t, while the foreign market does not, the $(r/w)_t$ at home exceeds (is less than) r/w abroad when $t < 1 (> 1)$, even with product prices equalized ([100] and [101]) Foreign factor prices are read off of the $t = 1$ curve in the left half of Fig 8 The possibility of value intensity reversals raises the even more interesting point made by Bhagwati and Srinivasan [23] that multiple equilibria are possible with differentials so that even if both product prices *and* differentials are identical in two countries, factor price equalization may not occur if one country is at one non-specialized equilibrium point and the other country at another

The effects of distortions on the pattern of trade in two-factor models are the same as those developed earlier for one-factor models trade can go either way so that the Heckscher–Ohlin model appears to break down Cases of perverse trade specialization have been mentioned either explicitly or implicitly by Bhagwati and Ramaswami [20], Caves [31], Haberler [52], Hagen [54], Johnson [77], Lewis [95], Magee [100] and [101], and Ohlin [118] among others

Following [101], consider a country without monopoly power in trade in which the two-factor intensities correspond (Batra and Pattanaik [11] extend [101] to permit variable international terms of trade) If the *export* industry pays a differential, both exports and imports decline toward zero, trade becomes small relative to national income, and a reversal of the pattern of trade is possible. With the export differential, the price of the physically abundant factor in the country falls in both industries below its price in foreign markets (even with free trade and internationally equalized product prices) so that the abundant factor in the distorted home market will wish to emigrate If permitted, this will speed up the trade reversal process, leading to eventual production specialization in what was formerly the imported product

On the other hand, if the *import-competing* industry pays a differential, both imports and exports increase, strengthening the original pattern of trade, and trade becomes larger relative to national income The price of the physically abundant factor in the country rises in both industries above its price abroad, resulting in a paradoxical flow of factors from regions where they are already scarce to the country where they are relatively abundant This moves the country toward production specialization in the initial export product Next, assume that capital is mobile internationally but labour is not and there are greater interindustry barriers to capital flows domestically than there are internationally If the *import-competing* industry pays a differential on capital, its non-intensive factor, and tries to offset this disadvantage by having the government impose a tariff on imports, then it is possible to have capital cross-flows, i e. capital leaving the export industry and going abroad while foreign capital is being attracted into the import-competing industry

If we denote by sets I and II whether the original pattern of trade has or has not been reversed, the Edgeworth–Bowley box may be divisible into as many as eight distinct regions [**102**]. Regardless of the definition of factor intensity used, four of these yield apparent empirical support for the Heckscher–Ohlin theorem (the observed relative factor intensity of the exported good corresponds to the relative factor abundance of the country) while the other four give Leontief-type results

We have already noted that in the normal case, an increase in the differential paid by an industry on either of its factors may raise the relative price of its product ([**11**], [**82**], and [**101**]). If the increase occurs in the export sector, the country's terms of trade will improve and if it occurs in the import-competing sector, they will deteriorate, given appropriate stability assumptions. Batra and Pattanaik [**11**] and Jones [**82**] note that a reversal of only one of the factor intensities can lead to the opposite result, namely, an increased export differential being associated with a decrease in the terms of trade.

Finally, a word on stability in the static models ([**68**] and [**70**] consider stability in dynamic contexts) Jones [**82**] notes that in the abnormal cases where the physical and value factor intensities do not correspond, changes in the differential, t, or the capital–labour factor endowment, k, elicit price responses which reinforce rather than moderate the initial change, so that the conditions required for stability 'are apt to be rather stringent' and that in the abnormal case, high values for S_x and S_y endanger stability rather than encourage it, contrary to the normal case

V. Differentials: the welfare effects

The welfare consequences of factor price differentials in both one- and

two-factor models have elicited a lot of interest. Manoilesco [104] starte one debate in 1931 with his observation that Eastern European countrie were being hurt by trade with Western Europe because free trade wa forcing them to specialize in slow growth and low productivity agricultural pursuits while Western Europe benefited from specialization in manufactured products, where productivity was higher The differentials in the marginal productivities in the two sectors convinced him that protection of the manufacturing industries in less developed countries would be superior to free trade This theme was formalized for both one- and two-factor models by Hagen [54] At least three propositions can be made on this issue

First, protection may or may not be superior to free trade with factor price differentials This traditional second-best argument has been made by a number of authors ([**20**], [**52**], and [**77**], among others)

Second, from a welfare point of view, protectionism is not the optimal policy response to differentials This proposition has had an interesting history of independent rediscovery over a thirty-year period. The proposition was made by Ohlin [**118**] in 1931, restated by Meade [**109**] and Hagen [**55**] in the 1950s and rediscovered by Bhagwati and Ramaswami [**20**] in 1963 The optimal policy is factor taxation or subsidy. To quote Ohlin · 'A cash bonus corresponding to the extra labour costs would bring the situation closer to a more normal one, than a system of duties can do' [**118**, p 44] Further, Ohlin noted that if the differential is caused by factor immobility, then duties will just increase the differential, whereas the rational policy is 'to increase the labour mobility and do away with the watertight labour compartment' [**118**, p 44]

Third, policies can be ranked in terms of their efficacy in offsetting differentials Bhagwati [**19**] ranks these policies as follows

1 First-best factor tax-cum-subsidy ,
2 Second-best production tax-cum-subsidy ;
3 Third-best tariff (trade subsidy) ,
4 Consumption tax-cum-subsidy will not help

A paper dealing with second-best policy intervention involving distortions without trade has been written by Fishlow and David [**42**].

In response to a paper by Kemp and Negishi [**86**], Bhagwati, Ramaswami, and Srinivasan [**21**] have written an extremely interesting paper in which they derive an equation, now quite in vogue, which yields some interesting results (this paper is a predecessor of [**19**]) Following [**20**], they consider the equality of the domestic rate of substitution (DRS), the domestic rate of transformation (DRT) and the foreign rate of transformation (FRT) They find that if only two out of three of these are equal

under *laissez-faire*, then there is some policy which will raise welfare, eve though the equality of the first two variables is broken thereby If a three of these variables are unequal, there is no feasible policy which wi raise welfare.

Bhagwati [19], following the theory of the second-best, states tha reduction of a differential with other distortions in the system may not increase welfare, while reduction of a differential with no other distortion in the system necessarily increases welfare. Bhagwati, Batra, and Pattanaik [11] and Foster and Sonnenschein [45] note that the second proposition is contingent upon the existence of a single equilibrium point Batra and Pattanaik show that an increased differential lowers real income, and hence welfare, in the normal cases and increases income and welfare in the abnormal cases with product prices fixed Thus, a decrease in one distortion, even with no other distortions present, does not necessarily mean an improvement in welfare (this is another consequence of distortion-induced reversals of the value factor intensities of the products which make multiple equilibria possible)

In another paper, Batra and Pattanaik [9] find that a *decrease* in a country's terms of trade may lead to an *increase* in welfare if factor markets are distorted This may have a bearing on the debate over the secular deterioration of the terms of trade of the less developed countries since their factor markets are notoriously imperfect [1] Empirical work should be undertaken to investigate the importance of this point Finally, Kemp and Negishi [87] examine welfare with distortions in a model with many products and factors and variable returns to scale

VI. Empirical studies

This section will review briefly some empirical work on factor price differentials This discussion is included as a supplement to the theoretical review, and for that reason is not an exhaustive treatment In Section I we reviewed several causes for both distortionary and non-distortionary differentials Excellent examples of the latter are given in Mincer's survey [112] The econometric, methodological, and measurement problems associated with measuring differentials have been investigated by Bahral [6], Hall [57], Hanna and Denison [58], Hart [63], Johansen [75], and Perlman [124] An outstanding paper with good coverage of the pre-1962 literature has been written by Reder [129] The studies by Clark [32, Ch 10] and Bellerby [14] are examples of some of the applied work in

[1] This argument was extended to the immiserizing growth case in R Batra and G W Scully, 'The theory of wage differentials welfare and immiserizing growth', *Journal of International Economics*, 1 (May 1971), pp 241–7.

the area In this section we will consider four topics. immobility, union power, the corporate income-tax, and factor distortions with trade.

The reluctance of factors to move in response to higher real incomes may involve a quantum effect they will not move to the higher income area or industry until the differential passes a certain threshold [**118**]. This inertia was demonstrated by Lester [**93**], who found in his survey of forty-eight labour market areas that the high-wage to low-wage ratio for identical jobs in the same community sometimes averaged 1·5 Meyers [**110**, p 94] noted that 'there is a good bit of evidence to show that differences between job and no job are more effective in inducing movement of labour than differences in wages'. Stafford [**145**] observes that unionization probably reduces geographical mobility (although Parker and Burton [**122**] were not able to attribute a pre-war to post-war decline in U S mobility to unionization) Two interesting studies on migration have attempted to calculate the wage differential required to induce migration ([**144**] and [**155**]). Sjaastad [**144**] finds in the U.S. that 'the typical migrant would be indifferent between two destinations, one of which was 146 miles more distant than the other, if the average annual labour earnings were $106 higher (1947–9 dollars) in the more distant one'. Vanderkamp [**155**] makes similar calculations for Canada, finding somewhat smaller income–distance trade-offs. Benham, Maurizi, and Reder [**15**], in a study of the mobility of physicians and dentists in the United States, found that the latter are much more impeded geographically than the former because of state licensing requirements

Marshall's work on the determinants of union power has been reviewed by Allen [**2**] These determinants include an inelastic demand for the product's output, technological necessity, a low proportion of total costs attributable to labour (i e a capital-intensive unionized sector) and inelastic supplies of the co-operating factors A large body of empirical work has been done on union/non-union wage differentials in the United States where roughly a fourth of the labour force is unionized [**94**, p 5]. Many of the earlier studies on unionization are discussed in Lewis [**94**], which is the standard reference work up to 1963 Several of these studies are summarized chronologically in Table 1. Roughly nine estimated that unions raise union over non-union wages, three concurred but with qualification depending on the phase of the business cycle, five felt unions had little or no effect, and Douglas [**37**] found that the union/non-union wage ratio was less than one and declining in the early twentieth century.

One of the most interesting results is that unions improve their position *vis-à-vis* non-union labour in troughs but are hurt at peaks of the business cycle (see [**94**], [**125**], and [**147**]) Thus, in addition to utilizing idle resources, high levels of employment also work to reduce distortions i

TABLE 1

A survey of the effects of unionism on wages in the U.S.

	Author	Date of study	Scope of study	Conclusions
1	Douglas [37]	1930	Real wages in US, 1890–1926	In 1920, the ratio of union/non-union wages was far below the level in 1890–9, ratio began to rise in early 1920s
2	Friedman [46]	1951	Comment on post-war inflation	The effect of unions on wages exaggerated, not over 10–20% of labour force affected by unions
3	Sultan [147]	1954	Measured ratio of share of union/non-union (1929 = 100) between 1929 and 1951	Unions showed relative improvement in 1931–4, 1934–8, 1941–9 and declined in 1929–31, 1939–40, 1949–51
4	Looks [98]	1955	Investigated Cleveland, Ohio, labour market, 1945–50	Found greater tendency for pattern-bargaining among union than non-unionized plants
5	Ulman [153]	1955	Comment on Friedman [46]	Unions may be able to affect the structure of wages without affecting the level of wages.
6	Eisemann [40]	1956	Inter-industry wage changes 1939–47	Does not find that unions affected wage increases significantly over this period
7.	Maher [103]	1956	Used BLS data for 7 industries in 1950 cross-section study	Found no significant differences in union and non-union wages
8	Simler [143]	1961	Data 1899–1954	Found little relationship between labour's share in manufacturing industries and unionism
9	Lurie [99]	1961	Data for transit industry for selected years, 1920–48	Union over non-union returns 1920s 15–20%, early 1930s 20–25%, late 1930s 5–10%, 1940s below 18%
10	Ozanne [120]	1962	Occupational differentials for McCormick Works (1858–1959)	Prosperity widens and depression narrows occupational differentials
11	Lewis [94]	1963	Unionism since the 1920s in the US	Union / non-union differentials highest in depressions and lowest at peaks of economic activity Union over non-union returns 1932–3 over 25% 1947–8 less than 5%
12	Segal [141]	1964	Case studies	Unions are more successful with industries servicing local markets than those in national markets and with concentrated rather than competitive industries
13	Weiss [161]	1966	Compared union and non-union private wage and salary income for 1959 reported in the 1/1000 sample of the 1960 Census	The union/non-union differential largely evaporates after personal characteristics of the labour force are considered

TABLE 1 (*cont.*)

Author	*Date of study*	*Scope of study*	*Conclusions*
14. Stafford [145]	1968	Used portions of 1966 Survey of Consumer Finances gathered by the Survey Research Center, University of Michigan, to modify Weiss [161]	The union/non-union differential is 26% for operatives, 18% for clerical and sales workers and 0% (not statistically significant) for skilled labour (professionals, non-self-employed managers, etc.), even after personal characteristics of the labour force are considered.
15. Clover [33]	1968	Used BLS data of 31 surveys of earnings in 23 manufacturing industries, 1960–5	Union wages exceeded non-union wages by 18% although the differential narrowed somewhat over the period
16. Pierson [125]	1968	Combines quarterly data for U.S. manufacturing (1953–66) into two groups union and non-union	Average annual wage changes union, 3.62% non-union, 2.97%, with higher profit rates in union strength closely tied to adaptability to cost-of-living changes. Union strength has less impact in periods of low unemployment
17. Throop [151]	1968	Cross-section data for 1950 and 1960 for 2-digit manufacturing and 7 other industries	Estimates that the union/non-union wage differential increased 12.7% points 1950–60, after allowing for skill-level changes, city size, and labour supply.
18. Ashenfelter [4]	1971	Used 1967 Survey of Economic Opportunity data for U.S. urban males (8,123 white and 3,897 black) to investigate union effects on wage discrimination	The union/non-union wage differential is 9.7% for white males and 20.5% for black males. Craft unions are more discriminatory than industrial unions. Black/white male wages were 3.4% higher in 1967 than they would have been in the absence of unionism

labour pricing caused by contractual rigidity in union wages. Fishlow and David [42, p. 534] argued, following Hagen [54], that the microeconomic 'direction of [a] differential will, in general, be determined by demand conditions.' To this sectoral effect we can add that they are also affected by changes in *aggregate* demand.

It is of some interest to compare the proportion of workers which are unionized by country. Table 2 is reproduced from Utsumi [154] and shows that the lowest percentage of the group is in the United States while the highest is in the United Kingdom and West Germany.

This section will not systematically review studies dealing with differentials in the returns to capital. Risk aversion is the most frequently mentioned cause (see, for example, Fisher's paper [41]). In an early study, Bain [7] found that average post-tax profit rates were 12·1 per cent in

concentrated industries and 6·9 per cent in all other industries It is not clear what portion of these monopoly-oligopoly profits reflect differential returns to capital Segal [**141**] and others have tried to establish a relationship between factor price differentials and industries which have product market power. Schwartzman [**139**] was unable to establish a relationship between monopoly and either high or low wages On the other hand, Weiss [**161**] found relatively higher wages in concentrated industries, but

TABLE 2

Country	*Degree of unionization*	*Year of measurement*
1. United States	27–31%	1953
2. Japan	38 9%	1955
3. France	40 0%	1950
4. West Germany	44 2%	1953
5. United Kingdom	44 6%	1953

SOURCE. Utsumi [**154**]

oted that this was due to the economic superiority of labour in these idustries (higher productivity, embodied human capital, etc)

There is a body of rather controversial literature dealing with the icidence of the U S corporate income-tax If it were a tax on pure rofits, there should be no attempted shifting through price, output, or ıctor market effects Hall [**56**] was unable to find evidence of shifting in ither the product or factor markets while Harberger [**60**] found that apital bears between 112 and 120 per cent of the tax burden, so that it istorts the factor market for capital between the corporate and non-orporate sectors The question is a complicated one, however, and the ariance in the two results just cited is symptomatic of the lack of con-ensus in this area

Finally, a number of studies have related differentials to international rade. A good general discussion of labour and trade is provided by litchell [**113**] Kravis [**92**] found that U S wages in 1947 in the leading xport industries were $1·46 while they were only $1 23 in import-ompeting industries. Keesing [**84**] found that United States exports are killed-labour intensive while import-competing products use relatively ırge amounts of unskilled labour in their production, which helps explain Kravis's earlier result (see also Mitchell [**114**]) Bourque [**25**] shows that mport-competing industries locate primarily in the South, where skill evels are lower than the national average Salant [**132**] has discussed rade and the corporate income-tax while Marx [**106**] has discovered that lifferentials penalize U.S. exporters relative to importers in the purchase

of non-primary inputs such as ocean freight. Several studies have examined inter-country wage differences: Mitchell [114] and Papola and Bharadwaj [121], to name two

To date, I know of no empirical work which explores the theoretical question of physical and value factor intensity reversals discussed in the previous sections. This is certainly an important and fruitful area for future research

VII. Concluding remarks

One is struck, in doing a survey, at the cyclical repetition of economic discoveries For its time, probably the most outstanding work out of the 162 papers surveyed is the book review written by Ohlin [118] in 1931. It is purely qualitative but it arrives at many of the modern results of the structural and welfare effects of differentials in the normal case The following quote from Ohlin [118] brilliantly anticipates the recent theoretical work on unionization the adverse output effects in the unionized industry and expansion of the non-unionized industry ([11], [23], [69], [71], [82], [97], [100], [101], and [115]), the decline in real income and welfare attributable to the differential ([9], [11], [19], [20], [21], [42], [52], [54], [77], [86], and [109]), the dependence of the reallocation of labour on the elasticity of demand for the final product, which is particularly high for internationally traded goods ([37], [60], [82], [100], [101], [111]), the dependence of the effects on the product factor intensities ([11], [23], [60], [69], [71], [78], [80], [82], [97], [100], [101], [111]), and the welfare superiority of factor subsidies to import duties ([8], [9], [10], [11], [19], [20], [21], [52], [77], and [109]) ·

> The high wage industry, where the marginal productivity of labour is relatively great, is kept back, while the low wage industry develops more than it would have done under conditions of uniform wage rates The national income in terms of goods and services is reduced. It depends upon the circumstances in each particular case, whether this influence on the distribution of labour between various industries is great or not. Generally, a forcing up of wages in an industry will have greater influence the more elastic is the demand for its products For this reason the pursuance of such a policy in export industries will reduce the number of workers there employed much more than its practice in home market industries. If the wages paid to workers in the trade union makes up only a small part of the commodity price [i e , capital intensive] the decline in sales and, thus, the shift in production is likely to be small . . . [For welfare purposes] a cash bonus corresponding to the extra labour costs would bring the situation more close to a 'normal' one, than a system of duties can do. [118, pp. 39-40, 44]

Obviously, book reviews today are just not what they used to be.

University of Chicago

REFERENCES

1. ADAMS, F. G., 'The size of individual incomes socioeconomic variables and chance variation', *Review of Economics and Statistics*, 40 (Nov 1958), pp. 390–8.
2. ALLEN, BRUCE T., 'Market concentration and wage increases, U.S. manufacturing, 1947–1964', *Industrial and Labor Relations Review*, 21 (Apr. 1968), pp 353–66.
3. ALTMAN, STUART H, and FISHER, ANTHONY C, 'Marginal product of labor, wages and disequilibrium. comment', *Review of Economics and Statistics*, 51 (Nov. 1969), pp. 485–6
4. ASHENFELTER, ORLEY, 'Racial discrimination and trade unionism', *Journal of Political Economy*, 80 (May, 1972)
5. AUERBACH, ROBERT, 'The effects of price supports on output and factor prices in agriculture', *Journal of Political Economy*, 78 (Nov./Dec. 1970), pp. 1355–61.
6. BAHRAL, URI, 'Wage differentials and specification bias in estimates of relative labor prices', *Review of Economics and Statistics*, 44 (Nov. 1962), pp. 473–81
7. BAIN, J. S., 'Relation of profit rate to industry concentration. American manufacturing, 1936–1940', *Quarterly Journal of Economics*, 65 (Aug. 1951), pp 293–324
8. BARDHAN, P. K, 'Factor market disequilibrium and the theory of protection', *Oxford Economic Papers*, 16 (Nov. 1964), pp. 375 88.
9. BATRA, R., and PATTANAIK, P. K, 'Domestic distortions and the gains from trade', *Economic Journal*, 80 (Sept. 1970), pp. 638–49.
10. ——— 'Factor market imperfections and gains from trade', *Oxford Economic Papers*, 23 (July 1971), pp 182–8
11. ——— 'Factor market imperfections, the terms of trade and welfare', *American Economic Review*, 61 (Dec 1971), pp 946–55
12. BEHMAN, SARA, 'Wage changes, institutions and relative factor prices in manufacturing', *Review of Economics and Statistics*, 51 (Aug. 1969), pp 227–38.
13. BELL, DURAN, 'Occupational discrimination as a source of income differences. lessons of the 1960's', *American Economic Review*, 62 (May, 1972), pp 363–72.
14. BELLERBY, J. R., in association with G. R. Allen and others, *Agriculture and Industry Relative Income* London Macmillan, 1956.
15. BENHAM, LEE; MAURIZI, ALEX, and REDER, MELVIN W, 'Migration, location and remuneration of medical personnel physicians and dentists', *Review of Economics and Statistics*, 50 (Aug. 1968), pp. 332–47 (reprinted in [29]).
16. BHAGWATI, JAGDISH, 'The pure theory of international trade a survey', *Economic Journal*, 74 (Mar. 1964), pp. 1–84.
17. —— *The Theory and Practice of Commercial Policy*, Frank Graham Memorial Lecture (1967), Special Papers in International Economics No. 8, Princeton University, 1968.
18. —— 'Distortions and immiserizing growth a generalization', *Review of Economic Studies*, 35 (Oct 1968), pp 481–5.
19. —— 'The generalized theory of distortions and welfare', in J. Bhagwati, R. Jones, R. Mundell, and J. Vanek, eds., *Trade, Balance of Payments and Growth*, Papers in International Economics in Honour of Charles P. Kindleberger Amsterdam: North Holland, 1971.
20. —— and RAMASWAMI, V. K., 'Domestic distortions, tariffs and the theory of optimum subsidy', *Journal of Political Economy*, 71 (Feb 1963), pp. 44–50.
21. ——— and SRINIVASAN, T. N., 'Domestic distortions, tariffs and the theory of optimum subsidy. some further results', *Journal of Political Economy*, 77 (Sept. 1969), pp. 1005–10.
22. —— and SRINIVASAN, T. N., 'Optimal policy intervention to achieve non-economic objectives', *Review of Economic Studies*, 36 (Jan. 1969), pp. 27–38.

23. BHAGWATI, JAGDISH and SRINIVASAN, T. N., 'The theory of wage differentials production response and factor price equalization', *Journal of International Economics*, 1 (Feb. 1971), pp. 19–35.
24. BLACK, J, 'Foreign trade and real wages', *Economic Journal*, 79 (Mar. 1969), pp. 184–5.
25. BOURQUE, P J., 'Geographic earnings differentials and foreign trade', *Review of Economics and Statistics*, 40 (May 1958), pp. 177–9.
26. BRONFENBRENNER, M, 'Wages in excess of marginal revenue product', *Southern Economic Journal*, 16 (Jan. 1950), pp 297–309.
27. —— 'Potential monopsony in labor markets', *Industrial and Labor Relations Review*, 9 (July 1956), pp 577–88
28. BRYCE, HERRINGTON, J., 'Regional labor earnings differentials in a small developing country the Republic of Panama', *Journal of Regional Science*, 9 (Dec 1969), pp 405–15
29. BURTON, JOHN F., jr., BENHAM, LEE K.; VAUGHN, WILLIAM M. III, and FLANAGAN, ROBERT J., *Readings in Labor Market Analysis*. New York: Holt, Rinehart and Winston, 1971.
30. CAIRNES, J. E., *Some Leading Principles of Political Economy*. London Macmillan, 1874.
31. CAVES, RICHARD, *Trade and Economic Structure*. Cambridge, Mass.: Harvard University Press, 1960, pp. 58–68.
32. CLARK, COLIN, *The Conditions of Economic Progress*. London: Macmillan, 1957.
33. CLOVER, VERNON T., 'Compensation in union and nonunion plants, 1960–1965', *Industrial and Labor Relations Review*, 21 (Jan. 1968), pp. 226–33.
34. CORDEN, W M., *Recent Developments in the Theory of International Trade*. International Finance Section, Department of Economics, Princeton University, 1965
35. —— 'Wage rigidity and the balance of payments: a third-best argument for tariffs?', mimeographed, 1971.
36. DESPRES, E, and KINDLEBERGER, C. P, 'The mechanism for adjustment in international payments— the lessons of postwar experience', *American Economic Review*, 42 (May 1952), pp 332–44
37. DOUGLAS, P H, *Real Wages in the United States, 1890–1926*. Houghton-Mifflin, 1930
38. EAGLEY, R. V., 'Market power as an intervening mechanism in Phillips curve analysis', *Economica, New Series*, 32 (Feb. 1965), pp. 48–64
39. ECKAUS, R. S, 'The factor proportions problem in underdeveloped areas', in A. N. Agarwala and S P Singh, eds., *The Economics of Underdevelopment*. New York: Oxford University Press, 1963, pp. 348–78 Reprinted from the *American Economic Review*, 45 (Sept. 1955), pp 539–65
40. EISEMANN, D M, 'Inter-industry wage changes, 1939–1947', *Review of Economics and Statistics*, 38 (Nov. 1956), pp 445–8.
41. FISHER, L, 'Determinants of risk premiums on corporate bonds', *Journal of Political Economy*, 67 (June 1959), pp. 217–37
42. FISHLOW, ALBERT, and DAVID, PAUL A, 'Optimal resource allocation in an imperfect market setting', *Journal of Political Economy*, 69 (Dec. 1961), pp. 529 46.
43. FOGEL, W. A, 'Job rate ranges a theoretical and empirical analysis', *Industrial and Labor Relations Review*, 17 (July 1964), pp. 584–97.
44. FORCHHEIMER, KARL, 'The role of relative wage differences in international trade', *Quarterly Journal of Economics*, 62 (Nov. 1947), pp. 1–30.
45. FOSTER, EDWARD, and SONNENSCHEIN, HUGO, 'Price distortion and economic welfare', *Econometrica*, 38 (Mar. 1970), pp. 281–97.
46. FRIEDMAN, MILTON, 'Some comments on the significance of labor unions for

economic policy', in David M. Wright, ed., *The Impact of the Union*. New York: Harcourt Brace, 1951, pp. 204–34.

47. Fuchs, Victor R., and Perlman, Richard, 'Recent trends in Southern wage differentials', *Review of Economics and Statistics*, 42 (Aug. 1960), pp 292–300.
48. Fukuchi, Takao, and Nobukuni, Makoto, 'An econometric analysis of national growth and regional income equality', *International Economic Review*, 11 (Feb. 1970), pp. 84–100.
49. Gallaway, Lowell, 'The North-South wage differential', *Review of Economics and Statistics*, 45 (Aug. 1963), pp. 264–72.
50. Garbarino, Joseph, 'A theory of interindustry wage structure variation', *Quarterly Journal of Economics*, 64 (May 1950), pp. 282–305.
51. Haberler, Gottfried, *The Theory of International Trade*. New York: Macmillan, 1937.
52. —— 'Some problems in the pure theory of international trade', *Economic Journal*, 60 (June 1950), pp 223–40.
53. —— 'Real cost, money cost and comparative advantage', *International Social Science Bulletin*, 3 (Spring, 1951), pp. 54–8
54. Hagen, Everett E, 'An economic justification of protectionism', *Quarterly Journal of Economics*, 72 (Nov. 1958), pp. 496–514
55. —— 'Reply', *Quarterly Journal of Economics*, 75 (Feb. 1961), pp. 145–51.
56. Hall, Challis A., 'Direct shifting of the corporation income tax in manufacturing', *American Economic Review*, 54 (May 1964), pp. 258–71.
57. Hall, Robert E, 'Why is the unemployment rate so high at full employment?', *Brookings Papers on Economic Activity* (No 3, 1970), pp 369–410.
58. Hanna, Frank A. (and E. F Denison's comment), 'Analysis of interstate income differentials theory and practice', in Conference on Research in Income and Wealth, *Regional Income* ('Studies in Income and Wealth', Vol. 21). Princeton, N.J Princeton University Press (for the National Bureau of Economic Research), 1957, pp. 113–79.
59. Hanoch, Giora, 'An economic analysis of earnings and schooling', *Journal of Human Resources*, 2 (No. 3, 1967), pp 310–29.
60. Harberger, Arnold C., 'The incidence of the corporation income tax', *Journal of Political Economy*, 70 (June, 1962), pp 215–40.
61. Harcourt, G. C., 'Some Cambridge controversies in the theory of capital', *Journal of Economic Literature*, 7 (June 1969), pp. 369–405
62. Harris, J. R, and Todaro, M. P., 'Migration, unemployment and development: a two-sector analysis', *American Economic Review*, 60 (Mar. 1970), pp 126–42.
63. Hart, P. E, 'Statistical measures of concentration vs. concentration ratios', *Review of Economics and Statistics*, 43 (Feb 1961), pp. 85–6.
64. Heckscher, Eli, 'The effect of foreign trade on the distribution of income', *Ekonomisk Tidskrift*, 21 (1919), pp 497–512. Reprinted in H S. Ellis and L. A. Metzler, eds., *Readings in the Theory of International Trade*. Homewood, Ill.: Richard D. Irwin, 1949, pp 272–300
65. Hemming, M. F. W., and Corden, W. M., 'Import restriction as an instrument of balance of payments policy', *Economic Journal*, 68 (Sept. 1958), pp 483–510.
66. Henderson, A., 'The restriction of foreign trade', *Manchester School of Economics and Social Studies*, 17 (Jan. 1949), pp. 12–35.
67. Henderson, John P., 'An intercity comparison of differentials in earnings and the city worker's cost of living', *Review of Economic Statistics*, 37 (Nov. 1955), pp. 407–11.
68. Herberg, Horst, 'On a two-sector model with non-shiftable capital and labour-market imperfections', *Zeitschrift für die gesamte Staatswissenschaft*, 128 (Apr. 1972), pp. 10–21.

69. HERBERG, HORST, and KEMP, MURRAY C., 'Factor market distortions, the shape of the locus of competitive outputs and the relation between product prices and equilibrium outputs', in J. Bhagwati, R. Jones, R. Mundell, and J. Vanek, eds., *Trade, Balance of Payments and Growth*, Papers in International Economics in Honour of Charles P. Kindleberger. Amsterdam: North Holland, 1971.

70. —— —— 'Growth and factor market "imperfections"', mimeographed, 1971

71. —— —— and MAGEE, STEPHEN P., 'Factor market distortions, the reversal of relative factor intensities, and the relation between product prices and equilibrium outputs', *Economic Record*, 47 (Dec. 1971), pp. 518–30

72. HIESER, R. O., 'Wage determination with bilateral monopoly in the labour market: a theoretical treatment', *Economic Record*, 46 (Mar. 1970), pp 55–72

73. HOLMES, R. A, and MUNRO, J. M., 'Regional nonfarm income differences in Canada: an econometric study', *Journal of Regional Science*, 10 (Apr 1970) pp. 65–74.

74. INADA, K., 'Investment in fixed capital and the stability of growth equilibrium', *Review of Economic Studies*, 33 (Jan. 1966), pp. 19–30.

75. JOHANSEN, LEIF, 'A note on the theory of interindustrial wage differentials', *Review of Economic Studies*, 25 (Feb 1958), pp. 109–13.

76. JOHNSON, D. G 'Labor mobility and agricultural adjustment', in E. O Heady, H. G. Diesslin, H. R Jensen, and G L. Johnson, eds, *Agricultural Adjustment Problems in a Growing Economy* Ames, Iowa Iowa State College Press, 1958.

77. JOHNSON, HARRY G, 'Optimal trade intervention in the presence of domestic distortions', in Baldwin, *et al*, *Trade, Growth and the Balance of Payments* Amsterdam: North Holland, 1965, pp 3–34.

78. —— 'Factor market distortions and the shape of the transformation curve', *Econometrica*, 34 (July 1966), pp. 686–98.

79. —— 'Minimum wage laws a general equilibrium analysis', *Canadian Journal of Economics*, 2 (Nov. 1969), pp. 599–604

80. —— and MIESZKOWSKI, PETER, 'The effects of unionization on the distribution of income a general equilibrium approach', *Quarterly Journal of Economics*, 84 (Nov 1970), pp 539–61

81. JONES, RONALD W., 'The structure of simple general equilibrium models', *Journal of Political Economy*, 73 (Dec 1965), pp. 557–72

82. —— 'Distortions in factor markets and the general equilibrium model of production', *Journal of Political Economy*, 79 (May/June 1971), pp 437–59.

83. KAFKA, A, '*A new argument for protectionism?*', *Quarterly Journal of Economics*, 76 (Feb. 1962), pp 163–6

84. KEESING, DONALD, 'Labor skills and comparative advantage', *American Economic Review*, 56 (May 1966), pp. 249–58.

85. KEMP, MURRAY C., 'Some issues in the analysis of trade gains', *Oxford Economic Papers*, 20 (July 1968), pp. 149–61.

86. —— and NEGISHI, T, 'Domestic distortions, tariffs and the theory of optimum subsidy', *Journal of Political Economy*, 77 (Nov. 1969), pp. 1011–13.

87. —— —— 'Variable returns to scale commodity taxes, factor market distortions and their implications for trade gains', *Swedish Journal of Economics*, 72 (June 1970), pp. 1–11.

88. KENEN, P. B., 'Development, mobility and the case for tariffs: a dissenting note', *Kyklos*, 16 (Fasc. 2, 1963), pp. 321–4.

89. KINDLEBERGER, CHARLES P., *International Economics*, 1st ed., Homewood, Ill.: Richard D. Irwin, 1953

90. —— *The Terms of Trade: A European Case Study*. Cambridge, Mass.. M.I T. Press, 1956.

91. KOO, A. Y. C., 'An economic justification for protectionism: comment', *Quarterly Journal of Economics*, 75 (Feb. 1961), pp. 133–44.

92. KRAVIS, IRVING B., 'Wages and foreign trade', *Review of Economics and Statistics*, 38 (Feb. 1956), pp. 14–30.

93. LESTER, RICHARD A., 'A range theory of wage differentials', *Industrial and Labor Relations Review*, 5 (July 1952), pp. 483–500.

94. LEWIS, H. G., *Unionism and Relative Wages in the United States*. Chicago. University of Chicago Press, 1963.

95. LEWIS, W. ARTHUR, 'Economic development with unlimited supplies of labour', in A. N. Agarwala and S. P. Singh, eds., *The Economics of Under-development*. New York. Oxford University Press, 1963, pp. 400–49. Reprinted from *The Manchester School*, 22 (May 1954), pp. 139–91.

96. LINDER, STAFFAN BURENSTAM, *An Essay on Trade and Transformation*. New York: John Wiley and Sons, 1961, pp. 76–8.

97. LLOYD, P. J., 'The shape of the transformation curve with and without factor market distortions', *Australian Economic Papers*, 9 (June 1970), pp 52–61

98. LOCKS, MITCHELL O., 'The influence of pattern-bargaining on manufacturing wages in the Cleveland, Ohio, labor market, 1945–1950', *Review of Economics and Statistics*, 37 (Feb. 1955), pp 70–6

99. LURIE, MELVIN, 'The effect of unionization on wages in the transit industry', *Journal of Political Economy*, 69 (Dec 1961), pp 558–72

100. MAGEE, STEPHEN P., 'Factor market distortions and the pure theory of international trade', unpublished Ph D. dissertation, Massachusetts Institute of Technology, May 1969.

101. —— 'Factor market distortions, production, distribution and the pure theory of international trade', *Quarterly Journal of Economics*, 75 (Nov 1971), pp 623–43 (a revision of Chapters 2 and 3 of [100])

102. —— 'Factor market distortions and empirical tests of the Heckscher–Ohlin theorem', Working Paper No 11 (Nov. 1971), Studies in International Business and Economics, Institute of International Studies, University of California, Berkeley (a revision of Chapter 4 of [100]).

103. MAHER, J E, 'Union, nonunion wage differentials', *American Economic Review*, 46 (June 1956), pp 336–52

104. MANOILESCO, MIHAIL, *The Theory of Protection and International Trade*. London. P. S King and Son, Ltd, 1931.

105. MANSFIELD, EDWIN, 'Wage differentials in the cotton textile industry, 1933–1952', *Review of Economics and Statistics* 37 (Feb. 1955), pp 77–82

106. MARX, DANIEL, 'Regulation of international liner shipping and "freedom of the seas"', *Journal of Industrial Economics* 16 (Nov. 1967), pp 46–62

107. MCGUIRE, TIMOTHY W., and RAPPING, LEONARD A 'The supply of labor and manufacturing wage determination in the United States an empirical examination', *International Economic Review*, 11 (June 1970), pp 258–68.

108. MCLEAN, A. A., 'Selective employment tax impact on prices and the balance of payments', *Scottish Journal of Political Economy*, 17 (Feb 1970), pp 1 17

109. MEADE, J. E, *Trade and Welfare* London Oxford University Press 1955

110. MEYERS, F, 'Price theory and union monopoly reply. *Industrial and Labor Relations Review*, 13 (Oct 1959), pp 94 5

111. MIESZKOWSKI, P M, 'On the theory of tax incidence', *Journal of Political Economy*, 75 (June 1967), pp 250 62

112. MINCER, JACOB, 'The distribution of labor incomes a survey with special reference to the human capital approach. *Journal of Economic Literature* 8 (Mar. 1970), pp 1–26.

113. MITCHELL, DANIEL J B., *Essays on Labor and International Trade* University of California at Los Angeles Institute of Industrial Relations, 1970

114. MITCHELL, EDWARD J., 'Explaining the international pattern of labor productivity and wages. a production model with two labor inputs' *Review of Economics and Statistics*, 50 (Nov. 1968), pp. 461–9.

115. MUNDLAK, YAIR, 'Further implications of distortion in the factor market', *Econometrica*, 38 (May 1970), pp. 517–32.

116. MYINT, H., 'Protection and economic development', in R. Harrod and D. C. Hague, eds, *International Trade Theory in a Developing World*. London Macmillan, 1963

117. —— *The Economics of Developing Countries* London: Hutchinson University Library, 1967.

118. OHLIN, BERTIL, 'Protection and non-competing groups', *Weltwirtschaftliches Archiv*, 33 (Heft 1, 1931), pp. 30–45.

119. —— *Interregional and International Trade*. Cambridge, Mass.: Harvard University Press, 1933

120. OZANNE, ROBERT, 'A century of occupational differentials in manufacturing', *Review of Economics and Statistics*, 44 (Aug. 1962), pp. 292–9.

121. PAPOLA, T S., and BHARADWAJ, V P, 'Dynamics of industrial wage structure. an inter-country analysis', *Economic Journal*, 80 (Mar. 1970), pp. 72–90.

122. PARKER, JOHN E., and BURTON, JOHN F., 'Volunteer labor mobility in the U.S. manufacturing sector', in Gerald G. Somers, ed, *Proceedings of the Twentieth Annual Winter Meeting* Industrial Relations Research Association, 1968, pp. 61–70 (reprinted in [29]).

123. PEITCHINIS, S. G, 'Occupational wage differentials in Canada, 1939–1965', *Australian Economic Papers*, 8 (June 1969), pp. 20–40.

124. PERLMAN, R, 'A note on the measurement of real wage differentials', *Review of Economics and Statistics*, 41 (May 1959), pp. 192–5.

125. PIERSON, GAIL, 'The effect of union strength on the U.S. "Phillips Curve" ', *American Economic Review*, 58 (June 1968), pp. 456–67.

126. RAMASWAMI, V K., 'International factor movement and the national advantage', *Economica*, 35 (Aug 1968), pp 305–27.

127. —— 'Welfare maximization when domestic factor movement entails external diseconomies', *Journal of Political Economy*, 78 (Sept./Oct. 1970), pp. 1061–8.

128. REDER, MELVIN W., 'The theory of occupational wage differentials', *American Economic Review*, 45 (Dec. 1955), pp. 833–52.

129. —— 'Wage differentials theory and measurement', in *Aspects of Labor Economics* New York National Bureau of Economic Research, 1962, pp. 257–99 (reprinted in [29]).

130. ROBINSON, JOAN, *The Economics of Imperfect Competition*. London Macmillan, 1934.

131. ROSS, A. M, and GOLDNER, W, 'Forces affecting the interindustry wage structure', *Quarterly Journal of Economics*, 64 (May 1950), pp. 254–81.

132. SALANT, WALTER S., 'The balance of payments deficit and the tax structure', *Review of Economics and Statistics*, 46 (May 1964), pp. 131–8.

133. SAMUELSON, M. C., 'The Australian case for protection re-examined', *Quarterly Journal of Economics*, 54 (Nov. 1939), pp. 143–9.

134. SAMUELSON, P A., *Foundations of Economic Analysis*. Cambridge, Mass: Harvard University Press, 1947, Chapter VIII

135. —— 'Theoretical notes on trade problems', *Review of Economics and Statistics*, 46 (May 1964), pp. 145–54.

136. SAWHNEY, P. K., 'Inter-industry wage differentials in India', *Indian Economic Journal*, 17 (Aug./Sept 1969), pp. 28–56.

137. SCHLESINGER, J. R, 'Market structure, union power and inflation', *Southern Economic Journal*, 24 (Jan 1958), pp. 296–312.

138. SCHWARTZMAN, DAVID, 'The effect of monopoly on price', *Journal of Political Economy*, 67 (Aug 1959), pp 352–62.

139. —— 'Monopoly and wages', *Canadian Journal of Economics and Political Science*, 26 (Aug 1960), pp 428–38.

140. SEGAL, M., 'Regional wage differences in manufacturing in the postwar period', *Review of Economics and Statistics*, 43 (May 1961), pp. 148 55.

141. —— 'The relation between union wage impact and market structure', *Quarterly Journal of Economics*, 78 (Feb. 1964), pp 96–114

142. SHEPHERD, WILLIAM G., 'Trends of concentration in American manufacturing industries, 1947–1958', *Review of Economics and Statistics*, 46 (May 1964), pp. 200–12.

143. SIMLER, N. J., 'Unionism and labor's share in manufacturing industries', *Review of Economics and Statistics*, 43 (Nov 1961), pp. 369–78.

144. SJAASTAD, LARRY A., 'The costs and returns of human migration', *Journal of Political Economy*, 70 (Oct 1962, supplement), pp. 80–93 (reprinted in [29])

145. STAFFORD, FRANK, 'Concentration and labor earnings. comment', *American Economic Review*, 58 (Mar 1968), pp. 174–81.

146. STOLPER, W , and SAMUELSON, P A , 'Protection and real wages', *The Review of Economic Studies*, 9 (Nov. 1941), pp 58–73.

147. SULTAN, PAUL E , 'Unionism and wage-income ratios 1929 1951', *Review of Economics and Statistics*, 36 (Feb. 1954), pp 67–73

148. TARSHIS, LORIE, 'Factor inputs and international price comparisons', in M. Abramovitz, *et al.*, eds., *The Allocation of Economic Resources* Stanford, Calif Stanford University Press, 1959, pp 236–44

149. TAUSSIG, F., *International Trade*. New York Macmillan, 1927, pp. 43–60.

150. TAYLOR, DAVID P , 'Discrimination and occupational wage differences in the market for unskilled labor', *Industrial and Labor Relations Review*, 21 (Apr. 1968), pp. 375–90 (reprinted in [29]).

151. THROOP, A. W , 'The union-nonunion wage differential and cost-push inflation', *American Economic Review*, 58 (Mar. 1968), pp 79–99

152. UEKAWA, Y., 'On a two-sector growth model with non-shiftable capital', mimeographed, 1970 (forthcoming in the *International Economic Review*).

153. ULMAN, L., 'Marshall and Friedman on union strength , *Review of Economics and Statistics*, 37 (Nov. 1955), pp. 384–401 (reprinted in [29]).

154. UTSUMI, Y., 'Rigidity of wage rate and the interfirm wage differentials', *Osaka Economic Papers*, 6 (Feb. 1958), pp 39–58.

155. VANDERKAMP, JOHN, 'Migration flows, their determinants and the effects of return migration', *Journal of Political Economy*, 79 (Sept./Oct 1971), pp 1012–31.

156. VINER, JACOB, 'The theory of protection and international trade a review', *Journal of Political Economy*, 40 (Feb 1932) pp 121 5

157. —— *International Trade and Economic Development* Oxford Oxford University Press, 1953

158. —— *Studies in the Theory of International Trade*. London George Allen and Unwin, reprinted, 1964, pp. 453–62, 493–500.

159. WEBB, L ROY, 'International factor movement and the national advantage. a comment', *Economica*, 37 (Feb 1970), pp. 81–4

160. WEINSTEIN, P A., 'Featherbedding a theoretical analysis', *Journal of Political Economy*, 68 (Aug. 1960), pp. 379–87

161. WEISS, LEONARD, 'Concentration and labor earnings , *American Economic Review*, 56 (Mar 1966), pp 96–117 (reprinted in [29])

162. WOLFSON, R. J , 'An econometric investigation of regional differentials in American agricultural wages', *Econometrica*, 26 (Apr 1958), pp 225–57

ON THE ELASTICITIES OF SUBSTITUTION AND COMPLEMENTARITY

By RYUZO SATO *and* TETSUNORI KOIZUMI[1]

1. Introduction and summary

ANY economist who has worked on the theories of production function, growth, and technical progress will no doubt realize how much he owes to Professor Hicks (and to Joan Robinson in fairness to her independent discovery) for his invention of the concept of the elasticity of substitution. His *Theory of Wages* has nourished more than a generation of economists since its first publication in 1932. It seems as though we have paid back our debt by numerous elaborations and extensions of the themes developed there Now Professor Hicks has again added another item to the debit side of our balance sheet the elasticity of 'complementarity' [4]

The concept of the elasticity of complementarity is not new, but what is novel is Hicks's discovery of an important missing link between Joan Robinson's definition of the old elasticity of substitution concept and his own definition Of course, the discovery would not have been made without extending the analysis to more than two-factor inputs, for the two definitions are equivalent under the two-factor constant returns to scale technology. Hicks came to realize the reciprocal nature of his 1932 definition when he considered the application of the concept to the so-called Marshall–Hicks rules of derived demand He found that the elasticity of derived demand λ is related to the Joan Robinson (and Allen partial) elasticity of substitution σ and the inverse of λ to the Hicks elasticity of 'complementarity' $c = 1/\sigma$ in a 'linear' fashion He further demonstrated that this redefinition of the old concept finds a more fruitful usage in the multifactor world of production (the three-factor case) in that the partial elasticity of complementarity is meaningfully defined as a dual concept to the Allen partial elasticity of substitution.

The duality relationship Hicks derived, however, was rather an awkward one from the formal point of view. It is the purpose of this paper to establish a more natural dual relationship between the partial elasticity of complementarity and that of substitution by appealing to the duality theory between production and cost functions. Interpreted in this manner, the introduction of the elasticity of complementarity concept, far from 'being too late for that',[2] sheds a new light on the theory of derived demand.

[1] This work was in part supported by the National Science Foundation Grant GS–3280. The authors wish to acknowledge Sir John Hicks for his very useful comments on an earlier draft

[2] Hicks [4], p. 296

After a brief discussion of the framework of analysis in the next section, Section 3 formally introduces the two partial elasticity concepts It is shown that the partial elasticity of complementarity measures the inverse of the cross elasticity of derived demand with marginal cost held constant, while the partial elasticity of substitution does the same thing with the level of output held constant. Section 4 establishes the formal relationship between the two concepts and indicates how factors can be characterized as substitutes and complements in a dual manner Another important dual result is established in Section 5 where it is shown that the partial elasticity of complementarity is negatively, and that of substitution positively, related to the elasticity of derived demand in the Marshall–Hicks formula Section 6 discusses the relationship of the partial elasticity concepts with two rather celebrated elasticity concepts—direct and shadow elasticities of substitution Finally, in Section 7, the use of these elasticities in analysing the behaviour of distributive shares is indicated

Two levels of discussion are employed in this paper The main line of arguments and the results will be first presented in non-technical terms Then the more technical derivations and detailed discussions are given in a starred section bearing the same number. The starred sections may be omitted by readers interested primarily in the results but should serve as useful references for more technically oriented readers

2. Duality between production and cost functions

The properties of demand functions for factor inputs are the main concern of the traditional theory of derived demand, where a typical question asked is 'what will be the effect on quantity demanded of a factor of a change in its own price, other prices, or output?' It is shown that the properties of the underlying production function play crucial roles in answering such a question and the factors are characterized as either substitutes or complements depending on how they enter the production process However, the properties of factor demand functions are equally well manifested in the corresponding cost function, which is a function of factor prices and output, in view of the fact that factor demand functions are 'implicitly' defined by the solution to the cost minimization problem The full implications of this correspondence between production and cost functions still need to be investigated for the further insight that could be shed on the theory of derived demand. We shall show in what follows how this duality theory can be applied in establishing the dual relationship between the elasticity of substitution and that of complementarity

The usefulness of the cost function in the factor demand theory is summarized by saying that the demand for, say, the ith factor input is

'explicitly' obtained by simply differentiating the cost function with respect to the price of this factor input. In most applications, especially for empirical estimation, the use of the cost function becomes a distinct advantage. This aspect of the problem, however, is outside the scope of the present paper

2*. *Consider a firm operating under the conditions of perfect competition in both product and factor markets. Let the firm's production function* be

$$x = f(a_1, \ldots, a_n) \tag{1}$$

where x = output and a_i = *ith* factor input We assume that f is a concave, twice continuously differentiable, and linear homogeneous function, strictly increasing in a, defined for $a \geqslant 0$. The rational behaviour of a firm deciding how much of each factor of production should be employed can be summarized in the minimization of the cost of production

$$C = \sum_{i=1}^{n} p_i a_i \tag{2}$$

subject to a given output prescribed by (1), where p_i = price of the *ith* factor input. The first-order conditions for the cost minimization require that

$$x = f(a),$$
$$p_i = \phi f_i \quad (1 \leqslant i \leqslant n), \tag{3}$$

where ϕ = Lagrange's multiplier which in equilibrium is equal to the marginal cost and $f_i = \partial f / \partial a_i$

The equilibrium conditions (3), provided that the second-order conditions are satisfied, define the demand functions for factor inputs

$$a_i = a_i(p_1, \ldots, p_n, x) \quad (1 \leqslant i \leqslant n). \tag{4}$$

Substituting these demand functions into (2), we obtain the cost function

$$C = g(p_1, \ldots, p_n, x), \tag{5}$$

which is a linear homogeneous, concave function in prices [1] This function is dual to the production function (1) in that the latter is linear homogeneous with respect to quantities of factor inputs and the former with respect to prices The usefulness of the cost function in the factor demand theory is summarized by the following relation:

$$a_i = g_i(p_1, \ldots, p_n, x) \quad (1 \leqslant i \leqslant n) \tag{6}$$

where $g_i = \partial g / \partial p_i$ [2] This implies that the properties of factor demand functions are incorporated in those of the cost function (5) as well In most applications, relation (6) turns out to be more useful than (4) in that (6) is an explicit relation while (4) is only implicitly defined by the equilibrium condition (3).

[1] The linear homogeneity of g with respect to prices is independent of the same condition on f. For a detailed discussion of the duality theory, see Diewert [2] and Shephard [10]

[2] The derivation of this result can be found, for example, in Diewert [2]

3. Partial elasticities of complementarity and substitution

Using the production and cost functions which are dual to each other, the partial elasticities of 'complementarity' and 'substitution' can be defined in a symmetric manner, the former reflecting the properties of the production function and the latter those of the cost function. The dual nature of these two partial elasticity concepts is best reflected in the roles they play in the comparative statics analysis. As for the Allen partial elasticity of substitution, it is well known that it registers the effect on the quantity demanded of one factor of a change in the price of another factor, where the partial derivative is taken holding output and other factor prices constant. The Hicks partial elasticity of complementarity, on the other hand, plays an exact dual role, namely, it registers the effect on the price of one factor of a change in the quantity of another factor, where the partial derivative is taken holding marginal cost and quantities of other factors constant.

3*. The *partial elasticity of complementarity* between factors a_i and a_j is defined as

$$c_{ij} = \frac{ff_{ij}}{f_i f_j}, \quad i \neq j, \tag{7a}$$

while the *partial elasticity of substitution* is defined as

$$\sigma_{ij} = \frac{gg_{ij}}{g_i g_j}, \quad i \neq j, \tag{7b}$$

where $f_{ij} = \partial^2 f/\partial a_i \partial a_j$ and $g_{ij} = \partial^2 g/\partial p_i \partial p_j$. To maintain the symmetry between the partial elasticity of complementarity c_{ij} and that of substitution σ_{ij}, it is more convenient to define c_{ij} as the inverse of Hicks's definition of s_{ij} [4, p. 291, equation (6)], i.e. $c_{ij} = 1/s_{ij}$. In this way, the duality between production and cost functions is more naturally incorporated and furthermore, as we shall show in Section 5 below, the 'linearity' relationship of the elasticity of derived demand with the partial elasticities is established in the multi-factor case as well. Both elasticities are symmetric in the sense $c_{ij} = c_{ji}$ and $\sigma_{ij} = \sigma_{ji}$, and have the properties that

$$\sum_{j=1}^{n} k_j c_{ij} = 0 \quad (1 \leq i \leq n) \tag{8a}$$

and

$$\sum_{j=1}^{n} k_j \sigma_{ij} = 0 \quad (1 \leq i \leq n), \tag{8b}$$

where

$$k_j = \frac{f_j a_j}{f} = \frac{g_j p_j}{g} = \text{relative share of factor } j. \tag{9}$$

The definition of the partial elasticity of substitution in (7b) differs from the original definition by Allen, who defined σ_{ij} in terms of the properties of the production function as

$$\sigma_{ij} = \frac{x}{a_i a_j} \frac{F_{ij}}{F} \tag{7'b}$$

where F is the bordered Hessian determinant of the production function and F_{ij} is the cofactor of f_{ij} in F. The two definitions are, of course, equivalent to each other, but the meaning of the concept becomes more transparent by interpreting (7'b).

Note that the bordered Hessian determinant F is associated with the comparative statics analysis of the equilibrium system (3) as

$$\begin{bmatrix} 0 & f_1 & . & . & . & f_n \\ f_1 & f_{11} & . & . & . & f_{1n} \\ . & & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & & & \\ f_n & f_{n1} & . & & . & f_{nn} \end{bmatrix} \begin{bmatrix} d\phi/\phi \\ da_1 \\ \\ . \\ . \\ da_n \end{bmatrix} = \begin{bmatrix} dx \\ dp_1/\phi \\ . \\ . \\ . \\ dp_n/\phi \end{bmatrix} \qquad (10b)$$

from which one obtains

$$\frac{\partial a_i}{\partial p_j} = \frac{1}{\phi}\frac{F_{ij}}{F}.$$

In elasticity terms, this becomes

$$\frac{\partial \log a_i}{\partial \log p_j} = \frac{f_j a_j}{f} \frac{f}{a_i a_j} \frac{F_{ij}}{F} = k_j \sigma_{ij}. \qquad (11b)$$

A dual result for the partial elasticity of complementarity can be derived as follows Consider the comparative statics analysis of (6) and $\phi = \phi(p_1, \ldots, p_n, x) = \text{constant}$. By taking the total differential of these equations, we get

$$\sum_{i=1}^{n} g_i \, dp_i = x \, d\phi,$$

$$g_i \frac{dx}{x} + \sum_{j=1}^{n} g_{ij} \, dp_j = da_i \quad (1 \leqslant i \leqslant n),$$

where use is made of the relations $\partial\phi/\partial x = 0$ and $\dfrac{\partial g_i}{\partial x}\dfrac{x}{g_i} = 1$ which follo from the linear homogeneity of f and g functions. Writing these as system, we have

$$\begin{bmatrix} 0 & g_1 & . & . & . & g_n \\ g_1 & g_{11} & . & . & . & g_{1n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ g_n & g_{n1} & . & . & . & g_{nn} \end{bmatrix} \begin{bmatrix} dx/x \\ dp_1 \\ \\ . \\ . \\ dp_n \end{bmatrix} = \begin{bmatrix} x \, d\phi \\ da_1 \\ . \\ \\ . \\ da_n \end{bmatrix} \qquad (1$$

from which one immediately obtains

$$\frac{\partial p_i}{\partial a_j} = \frac{G_{ij}}{G}.$$

In elasticity terms, this becomes

$$\frac{\partial \log p_i}{\partial \log a_j} = \frac{a_j p_j}{g} \frac{g}{p_i p_j} \frac{G_{ij}}{G} = k_j c_{ij}, \tag{11a}$$

where G is the bordered Hessian determinant of the cost function and G_{ij} is the cofactor of g_{ij} in G.

Duality theory outlined in Section 2 above enables us to derive the results in (11*a*) and (11*b*) in a much more straightforward manner. Going back to the equilibrium relation $p_i = \phi f_i$ (equation (3)), if the marginal cost is held constant, we get

$$\frac{\partial p_i}{\partial a_j} = \phi f_{ij} = \frac{p_i}{f_i} f_{ij}$$

This can be rewritten in elasticity form as

$$\frac{\partial \log p_i}{\partial \log a_j} = \frac{f_j a_j}{f} \frac{f f_{ij}}{f_i f_j} = k_j c_{ij},$$

which is equation (11*a*). Similarly, referring to the equilibrium relation $a_i = g_i$, if the level of output is held constant, we have

$$\frac{\partial a_i}{\partial p_j} = g_{ij}.$$

This can be converted into elasticity form as

$$\frac{\partial \log a_i}{\partial \log p_j} = \frac{g_j a_j}{g} \frac{g g_{ij}}{g_i g_j} = k_j \sigma_{ij},$$

which is equation (11*b*)

4. Substitutes and complements

As is well known, if there are only two factor inputs and if the production isoquants are convex to the origin, the two factors must necessarily be substitutes Only by increasing the number of factor inputs to more than two do we open up room for complementarity. However, the introduction of a new concept of the partial elasticity of complementarity necessitates the re-examination of the traditional treatment of the problem of substitutes and complements.

First, recall Allen's discussion with respect to the partial elasticity of substitution: two factor inputs are classified as substitutes or complements depending on whether the partial elasticity of substitution between the two inputs is positive or negative. Furthermore, among $n(n-1)/2$ total elasticities, at least $n-1$ of them must be positive, i e. there must exist at least $n-1$ 'substitutes' relationships. Can we make similar arguments with respect to the partial elasticity of complementarity?

In defining the partial elasticity of complementarity, it is to be noted that the level of output is not held constant. In fact, what this concept

measures is exactly the degree to which two factor inputs jointly contribute to a change in output, as the expression involves the cross partial derivative of the production function with respect to the two-factor inputs concerned Thus, if the partial elasticity of complementarity between, say, a_i and a_j is positive, it is reasonable to call these factors 'complements' in the sense that they work together to an increase in output level. Reversing the argument, the two factors may be called 'substitutes' when the partial elasticity of complementarity between the two is negative By following a similar argument as Allen, it is easy to show that among $n(n-1)/2$ total elasticities, at least $n-1$ of them must be positive.

The above criterion, although it makes good economic sense, is a bit confusing to say the least, for we now have two different concepts of substitutes and complements. This apparent difficulty, however, can be reconciled by following Hicks's suggestion. Recalling that the partial elasticity of substitution is related to the cross elasticity of derived demand when 'price' is changed, and the partial elasticity of complementarity to the inverse of the same elasticity when 'quantity' is changed, we may classify two factors as 'p-substitutes' or p-complements' depending on whether the partial elasticity of substitution between them is positive or negative, and 'q-complements' or 'q-substitutes' depending on whether the partial elasticity of complementarity is positive or negative When there are only two factor inputs, the two factors are thus necessarily 'q-complements' and 'p-substitutes'. However, a similar dual relationship does not in general carry over to the multi-factor case.

4*. The two bordered Hessian determinants F and G can be converted into the determinants of what may be called the partial elasticities of complementarity and substitution That is,

$$H = \begin{vmatrix} 0 & 1 & . & . & . & 1 \\ 1 & c_{11} & . & . & . & c_{1n} \\ . & & . & & . & \\ & . & . & . & . & \\ . & . & . & . & . & . \\ 1 & c_{n1} & . & . & . & c_{nn} \end{vmatrix} \tag{12a}$$

is defined as the *determinant of partial elasticities of complementarity* and

$$A = \begin{vmatrix} 0 & 1 & . & . & . & 1 \\ 1 & \sigma_{11} & . & . & . & \sigma_{1n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ & . & . & . & . & . \\ 1 & \sigma_{n1} & . & . & . & \sigma_{nn} \end{vmatrix} \tag{12b}$$

as the *determinant of the partial elasticities of substitution.*

Using the determinants of partial elasticities just defined, we can now rewrite the definitions (7′a) and (7′b) as

$$c_{ij} = \frac{1}{k_i k_j}\frac{A_{ij}}{A}, \tag{13a}$$

$$\sigma_{ij} = \frac{1}{k_i k_j}\frac{H_{ij}}{H}. \tag{13b}$$

The relation (13) defines the precise meaning of the statement that one partial elasticity is the inverse of the other[1] For example, in (13a), A_{ij}/A ($= A_{ji}/A$ since A is symmetric) is the element in the $i+1$th row and $j+1$th column of the matrix $[A]^{-1}$ Thus, the partial elasticity of complementarity between a_i and a_j corresponds to the 'inverse' element of the matrix of the partial elasticities of substitution, and the partial elasticity of substitution to that of the partial elasticities of complementarity

Recalling that the Allen partial elasticity of substitution was used to characterize factors as substitutes and complements, we may equally well use the Hicks partial elasticity of complementarity for that purpose, the latter being the definition with respect to changes in *quantities* while the former is with respect to *prices* of factor inputs Thus, following Hicks [4, p. 294], two factors a_i and a_j are defined as

$$\begin{Bmatrix} q\text{-complements} \\ q\text{-substitutes} \end{Bmatrix} \quad \text{according as } c_{ij} \gtrless 0$$

and

$$\begin{Bmatrix} p\text{-substitutes} \\ p\text{-complements} \end{Bmatrix} \quad \text{according as } \sigma_{ij} \gtrless 0.$$

From (8a) and (8b) and from the second-order conditions, we can conclude that *among $n(n-1)/2$ partial elasticities, there must be at least $n-1$ p-substitutes and $n-1$ q-complements* In particular, when there are only two factor inputs, they must necessarily be p-substitutes and q-complements. This correspondence is, of course, easy to see since we can directly evaluate (13a) or (13b) to get $\sigma_{12}\ c_{12} = 1$ However, a similar correspondence does *not* hold in the multi-factor case, except for some special types of functions such as Cobb–Douglas and CES

5. The partial elasticities and the market elasticity of derived demand: an extension of the Marshall–Hicks formula

Referring to the celebrated Marshall–Hicks formula, Hicks discovered an interesting 'linear' relationship between the market elasticity of derived demand and the partial elasticities of substitution and complementarity. Of course, in the two-factor case, this result is easy to verify since the

[1] For $n = 2$, both (13a) and (13b) reduce to $\sigma_{12}\ c_{12} = 1$ For $n = 3$, they reduce to equations (9) and (10) derived by Hicks [4, p 292] after eliminating 'own' elasticities c_{ii} and σ_{ii} by (8a) and (8b)

formula can be explicitly obtained. As it turns out, this monotone relationship can indeed be generalized to the multi-factor case.

For this purpose, the Marshall–Hicks formula must first be extended to the n-factor case. Since we have two partial elasticity concepts, the formula can now be expressed in dual forms Although the explicit solution is intractable, the above-mentioned monotone relationship can be verified indirectly by differentiation. Thus, it can be shown that *the elasticity of derived demand is related negatively to the partial elasticities of complementarity and positively to the partial elasticities of substitution.* That is, the greater is the degree of complementarity between any pair of factors the smaller will be the elasticity of derived demand, and the greater is the degree of substitutability between any pair of factors the greater will be the elasticity of derived demand

5*. To derive the Marshall–Hicks formula in the n-factor case, we work with the dual equilibrium conditions (equations (6) and (3)). First, define the following elasticities·

$$\eta = -\frac{dx}{dp}.\frac{p}{x} = \text{elasticity of demand for the product,}$$

$$\lambda = -\frac{da_1}{dp_1}.\frac{p_1}{a_1} = \text{elasticity of derived demand for the first factor,}$$

$$e_i = \frac{da_i}{dp_i}\frac{p_i}{a_i} = \text{elasticity of supply of other factors } (2 \leqslant i \leqslant n)$$

Substitute these elasticities into the total differential of (6) and (3) to get

$$-\frac{1}{\eta}\mathrm{d}\log x + \sum_{j=1}^{n} k_j c_{1j}\,\mathrm{d}\log a_j + \frac{1}{\lambda}\mathrm{d}\log a_1 = 0,$$

$$-\frac{1}{\eta}\mathrm{d}\log x + \sum_{j=1}^{n} k_j c_{ij}\,\mathrm{d}\log a_j - \frac{1}{e_i}\mathrm{d}\log a_i = 0 \quad (2 \leqslant i \leqslant n) \qquad (14a)$$

and

$$-\eta\,\mathrm{d}\log p + \sum_{j=1}^{n} k_j \sigma_{1j}\,\mathrm{d}\log p_j + \lambda\,\mathrm{d}\log p_1 = 0,$$

$$-\eta\,\mathrm{d}\log p + \sum_{j=1}^{n} k_j \sigma_{ij}\,\mathrm{d}\log p_j - e_i\,\mathrm{d}\log p_i = 0 \qquad (2 \leqslant i \leqslant n), \quad (14b)$$

where in (3) ϕ (= marginal cost) is replaced by p (= product price) since they are equal in equilibrium. ($14a$) and ($14b$) are systems of n equations in $n+1$ unknowns, unknowns being the rates of changes in $n+1$ quantity variables in ($14a$) and those in $n+1$ prices in ($14b$). To make the systems consistent, we can make use of the logarithmic differential of the cost and production functions

$$\mathrm{d}\log g = \sum_{j=1}^{n} k_j\,\mathrm{d}\log p_j + \mathrm{d}\log x$$

and

$$\mathrm{d}\log x = \sum_{j=1}^{n} k_j\,\mathrm{d}\log a_j,$$

to eliminate $\mathrm{d}\log x$ from (14*a*) and $\mathrm{d}\log p$ from (14*b*) respectively. Alternatively, we may treat these equations as additional equations yielding the following systems of $n+1$ equations in $n+1$ unknowns.

$$\begin{bmatrix} -1 & k_1 & k_2 & & k_n \\ -\frac{1}{\eta} & k_1 c_{11}+\frac{1}{\lambda} & k_2 c_{12} & \cdot\ \cdot & k_n c_{1n} \\ -\frac{1}{\eta} & k_1 c_{21} & k_2 c_{22}-\frac{1}{e_2} & \cdot\ \cdot & k_n c_{2n} \\ \vdots & \vdots & \cdot & & \\ -\frac{1}{\eta} & k_1 c_{n1} & k_2 c_{n2} & \cdot\ \cdot & k_n c_{nn}-\frac{1}{e_n} \end{bmatrix} \begin{bmatrix} \mathrm{d}\log x \\ \mathrm{d}\log a_1 \\ \\ \cdot \\ \\ \mathrm{d}\log a_n \end{bmatrix} = 0, \tag{15a}$$

$$\begin{bmatrix} -1 & k_1 & k_2 & & k_n \\ -\eta & k_1 \sigma_{11}+\lambda & k_2 \sigma_{12} & \cdot & k_n \sigma_{1n} \\ -\eta & k_1 \sigma_{21} & k_2 \sigma_{22}-e_2 & & k_n \sigma_{2n} \\ \vdots & \vdots & & \cdot & \vdots \\ -\eta & k_1 \sigma_{n1} & k_2 \sigma_{n2} & \cdot & k_n \sigma_{nn}-e_n \end{bmatrix} \begin{bmatrix} \mathrm{d}\log p \\ \mathrm{d}\log p_1 \\ \\ \vdots \\ \mathrm{d}\log p_n \end{bmatrix} = 0. \tag{15b}$$

Since (15*a*) and (15*b*) are homogeneous systems, we require that the determinants of coefficient matrices be vanishing for nontrivial solutions. These determinants can be converted to neater forms as

$$D = \det\left\{\begin{bmatrix} 0 & 1 \\ 1 & c_{ij} \end{bmatrix} + \begin{bmatrix} \eta & & & & 0 \\ & \frac{1}{k_1 \lambda} & & & \\ & & -\frac{1}{k_2 e_2} & & \\ & & & \cdot & \\ 0 & & & & -\frac{1}{k_n e_n} \end{bmatrix}\right\} = 0, \tag{16a}$$

$$\Delta = \det\left\{\begin{bmatrix} 0 & 1 \\ 1 & \sigma_{ij} \end{bmatrix} + \begin{bmatrix} \frac{1}{\eta} & & & & 0 \\ & \frac{\lambda}{k_1} & & & \\ & & -\frac{e_2}{k_2} & & \\ & & & \cdot & \\ 0 & & & & -\frac{e_n}{k_n} \end{bmatrix}\right\} = 0 \tag{16b}$$

To verify the monotone relationship between λ and σ_{ij}, for example, we first obtain

$$\frac{\partial\lambda}{\partial\sigma_{ij}} = \frac{k_1}{k_i k_j \Delta_{11}}[k_j^2\Delta_{ii} - 2k_i k_j \Delta_{ij} + k_i^2 \Delta_{jj}]$$
$$= \frac{k_1}{k_i k_j}\left[\frac{\Delta_{ii}}{\Delta_{11}}\left(k_j - k_i\frac{\Delta_{ij}}{\Delta_{ii}}\right)^2 + k_i^2\left(\frac{\Delta_{ii}\Delta_{jj}-\Delta_{ij}}{\Delta_{11}\Delta_{ii}}\right)^2\right]$$
$$(1 \leqslant i,j \leqslant n;\ i \neq j) \qquad (17)$$

By the well-known Jacobi theorem on determinants, we have

$$\Delta_{ii}\Delta_{jj} - \Delta_{ij}^2 = 0 \quad (0 \leqslant i,j \leqslant n).$$

Applying this theorem to (17), it is easy to see that $\partial\lambda/\partial\sigma_{ij} > 0$.[1]

By a similar argument, we can show that $\partial\lambda/\partial c_{ij} < 0$ Thus, we have proved

$$\frac{\partial\lambda}{\partial c_{ij}} < 0 \quad (1 \leqslant i,j \leqslant n,\ i \neq j), \qquad (18a)$$

$$\frac{\partial\lambda}{\partial\sigma_{ij}} > 0 \quad (1 \leqslant i,j \leqslant n,\ i \neq j) \qquad (18b)$$

6. A systematic formulation of the relationships among various concepts of the elasticities of substitution and complementarity

The previous sections have established the dual relationships between the partial elasticities of substitution and complementarity in the theory of derived demand We shall turn to the investigation of direct relationships among various elasticity concepts, for there are at least two more rather celebrated concepts direct elasticity and shadow elasticity in the multi-factor world of production

First, the direct elasticity of substitution d_{ij} between factors a_i and a_j is defined as the ratio between a percentage change of the factor proportion and a percentage change of the marginal rate of substitution given all other factors It can be easily seen that the inverse of d_{ij} is related to the determinant of the partial elasticities of *complementarity*.

Secondly, the shadow elasticity of substitution t_{ij} is defined in the same manner as the direct elasticity of substitution, only this time referring to the cost function rather than the production function And hence it is now related to the determinant of the partial elasticities of *substitution*.

All information concerning the nature of production technology is incorporated in the determinants of partial elasticities of complementarity

[1] A complete discussion of the Marshall rules for (15b) is found in Sato and Koizumi [8] The reader can verify $\partial\lambda/\partial\eta > 0$, $\partial\lambda e/\partial_j > 0$ $(2 \leqslant j \leqslant n)$ in (16a) as well With respect to the third rule, similar theorems can be established, *mutatis mutandis*

and substitution. The direct elasticity concepts then register those parts of that information relating directly to factors a_i and a_j, in the production function, while the shadow elasticity concepts register those in the cost function. More specifically it is easy to show that *the inverse of the direct (shadow) elasticity is a weighted average of all the proper partial elasticities of complementarity (substitution) related to factors i and j.*

6*. First, take the direct elasticity of substitution (DES) DES between factors a_i and a_j is defined as

$$d_{ij} = \frac{\dfrac{1}{a_i f_i} + \dfrac{1}{a_j f_j}}{-\dfrac{f_{ii}}{f_i^2} + 2\dfrac{f_{ij}}{f_i f_j} - \dfrac{f_{jj}}{f_j^2}}, \quad i \neq j \tag{19a}$$

Using the definition (7*a*) of the partial elasticity of complementarity, d_{ij} can be related to c_{ij}'s as

$$d_{ij}^{-1} = \frac{k_i k_j}{k_i + k_j} \begin{vmatrix} 0 & 1 & 1 \\ 1 & c_{ii} & c_{ij} \\ 1 & c_{ji} & c_{jj} \end{vmatrix}, \quad i \neq j. \tag{20a}$$

Thus, the inverse of DES is related to a minor of the determinant of the partial elasticities of complementarity.

The shadow elasticity of substitution (SES) is defined as

$$t_{ij} = \frac{-\dfrac{g_{ii}}{g_i^2} + 2\dfrac{g_{ij}}{g_i g_j} - \dfrac{g_{jj}}{g_j^2}}{\dfrac{1}{p_i g_i} + \dfrac{1}{p_j g_j}}, \quad i \neq j, \tag{19b}$$

where g is the cost function defined in (5). Making use of equation (7*b*) SES can be related to the partial elasticities of substitution as

$$t_{ij} = \frac{k_i k_j}{k_i + k_j} \begin{vmatrix} 0 & 1 & 1 \\ 1 & \sigma_{ii} & \sigma_{ij} \\ 1 & \sigma_{ji} & \sigma_{jj} \end{vmatrix}, \quad i \neq j \tag{20b}$$

The minors in (20*a*) and (20*b*) can be obtained from the determinants of partial elasticities of complementarity and substitution by eliminating all rows and columns except those which are related to factors a_i and a_j. If we eliminate own elasticities (elasticities with $i = j$), we obtain a more transparent relation. That is,

$$\frac{1}{d_{ij}} = (k_i + k_j)c_{ij} + \frac{k_j}{k_i + k_j} \sum_{\alpha \neq i,j} k_\alpha c_{i\alpha} + \frac{k_i}{k_i + k_j} \sum_{\beta \neq i,j} k_\beta c_{j\beta} \tag{21a}$$

and

$$t_{ij} = (k_i + k_j)\sigma_{ij} + \frac{k_j}{k_i + k_j} \sum_{\alpha \neq i,j} k_\alpha \sigma_{i\alpha} + \frac{k_i}{k_i + k_j} \sum_{\beta \neq i,j} k_\beta \sigma_{j\beta} \tag{21b}$$

Since the weights attached to the partial elasticities add up to unity, these elasticities are weighted averages of the proper partial elasticities of complementarity (or substitution).

7. Elasticities of complementarity and substitution and the behaviour of distributive shares

To conclude the paper, it would be useful to touch on a problem which has continuously attracted economists' attention: the relation between the elasticity of substitution and the behaviour of distributive shares. Hicks's partial elasticity of complementarity concept proves to be quite useful in this area as well. Since the authors have given an extensive argument of the subject elsewhere (Sato and Koizumi [9]), we shall merely report the results. The results can be stated as: *the relative share of one factor increases or decreases as the quantity (price) of another factor increases depending on whether the Hicks (Allen) partial elasticity of complementarity (substitution) between the two factors in question is greater or smaller than unity.*

Brown University
Ohio State University

REFERENCES

1. Allen, R. G. D., *Mathematical Analysis for Economists*, London: Macmillan, 1938.
2. Diewert, E. E., 'An application of the Shephard duality theorem: a generalized Leontief production function', *Journal of Political Economy*, 481–507, 1971.
3. Hicks, J. R., *The Theory of Wages*, 2nd edn., London: Macmillan, 1964.
4. —— 'Elasticity of substitution again: substitutes and complements', *Oxford Economic Papers*, 22, Nov. 1970, 289–96.
5. Marshall, A., *Principles of Economics*, 8th edn., New York: Macmillan, 1950.
6. Robinson, Joan, *Economics of Imperfect Competition*, London, 1933.
7. Samuelson, P. A., 'Two generalizations of the elasticity of substitution', in *Value, Capital, and Growth: Essays in Honour of Sir John Hicks*, ed. by J. N. Wolfe, Chicago: Aldine, 1968.
8. Sato, R., and Koizumi, T., 'Substitutability, complementarity, and the theory of derived demand', *Review of Economic Studies*, 27, Jan. 1970, 107–18.
9. —— —— 'The production function and the theory of distributive shares', *American Economic Review*, June 1973.
10. Shephard, R. W., *Theory of Cost and Production Functions*, Princeton, 1970.

THE DEATH-RATE OF 'TRACTORS' AND THE RATE OF DEPRECIATION

By CHARLES KENNEDY

I

IN *Capital and Growth* Sir John Hicks identified the rate of depreciation of a capital good with the proportion of the stock that had to be made good during the period.[1] In a footnote,[2] he claimed that this did not imply a confusion between depreciation in value terms and physical using up, or involve us in a concept of 'evaporation', in the manner of Professor Meade When I came to discuss Hicks's system in my contribution to *Value, Capital and Growth*,[3] I must have overlooked this footnote, because in making the same identification I thought it prudent to assume 'evaporation' explicitly; and I asserted, also in a footnote,[4] that the assumption of 'sudden death' would disturb the duality of the price and quantity equations.

I now think that we were both right and both wrong Hicks was right in thinking that a 'sudden death' assumption would not preclude the specification of the rate of depreciation as a given proportion of the original cost of the capital good, but wrong in identifying this rate with the death-rate of the capital good I was right in distinguishing between the death-rate and the rate of depreciation, but wrong in thinking that the 'sudden death' assumption disturbed the duality of the price and quantity equations As it turns out, the rate of depreciation depends on the rate of interest in *exactly the same way* as the death-rate depends on the rate of growth of the system Just as the rate of interest can be regarded as the 'dual' of the rate of growth, so the rate of depreciation can be regarded as the 'dual' of the death-rate [5] But the two rates will only be the same in the special limiting case in which the rate of interest is equal to the rate of growth All this is of course in the context of a regularly growing economy.

II

The demonstration is quite simple, and it will be convenient to take the death-rate first. We assume that the economy grows at a constant rate g We assume that our 'tractors' are scrapped at the end of n 'years', but that until they are scrapped every tractor is just as efficient as every other, i.e.

[1] [1], pp 161 et seq This identification is perhaps not quite explicit in Hicks's text, but it is certainly implied by his use of the same symbol in the price and quantity equations

[2] Ibid , p 161, *n* 3 [3] [2] [4] Ibid , p 289, n 9

[5] I use the term 'rate of interest' rather than 'rate of profit' because, as explained in [2], in period analysis we have strictly to draw a distinction between the rate of interest and the rate of profit on existing capital, and it is the rate of interest that is the 'dual' of the rate of growth.

the efficiency of a tractor does not depend on its age. This is all we need in order to work out the death-rate of tractors.

The constant rate of growth means that for every tractor in its last year of operation there will be $(1+g)$ tractors in their penultimate year, $(1+g)^2$ tractors in their ante-penultimate year, and so on until we reach the end of the sequence with $(1+g)^{n-1}$ tractors in their first year The death-rate (m, for mortality) will clearly be equal to the proportion of tractors in their final year Thus, evidently,

$$m = \frac{g}{(1+g)^n-1}. \tag{1}$$

To work out the rate of depreciation, we assume a constant rate of interest (i) In order to depreciate a tractor, we must lay aside at the end of each year of its life a constant sum such that the whole annuity accumulating at compound interest will be worth the price of a tractor at the end of the n years.[1] The rate of depreciation (d) will then be equal to the ratio of the annual charge (a) to the price of the tractor (p) Hence

$$d = a/p \tag{2}$$

where
$$p = a[1+(1+i)+(1+i)^2+\ldots+(1+i)^{n-1}], \tag{3}$$

so that
$$d = \frac{i}{(1+i)^n-1}. \tag{4}$$[2]

Thus, the duality of m and d is established,[3] but the two rates will only be the same in the special case where $g = i$ As is well known, this is the limiting case in which, with a 'classical' savings function, all profits are saved In the more general case, where some profits are consumed, i will be greater than g Since (m, d) are decreasing functions of (g, i), it follows that the rate of depreciation will normally be less than the death-rate.

III

I have called the ratio of the annual charge to the price of the tractor the 'rate of depreciation', although more strictly it should be called the 'rate of provision for depreciation', since this is what it measures rather than the rate at which the tractor actually depreciates in value. This latter rate will in fact increase as the tractor gets older. The absolute reduction

[1] There would appear to be some consensus that this 'fixed annuity' method of providing for depreciation is the theoretically 'correct' one in conditions of 'sudden death' Thus Professor Meade in [3], Chapter 8, calls it the 'rational' method Mr Sraffa in [5], pp 63 et seq , also insists that a constant annual charge is required. Cf also Joan Robinson [4]

[2] Sraffa's formula for the annual charge differs from that in (4) above because the former covers interest as well as depreciation If one subtracts the interest rate from Sraffa's formula, one obtains the formula in (4) I am indebted to Mr Sraffa for having made this clear to me in a comment on an earlier draft

[3] There is no doubt that this result was already implicit in, for example, Meade's discussion of depreciation in [3], Chapter 8 and Appendix III Meade was more concerned with other aspects of depreciation, however, and his treatment did not bring out the duality explicitly

in value will be a in the first year, $a(1+i)$ in the second year, $a(1+i)^2$ in the third year, and so on until it becomes $a(1+i)^{n-1}$ in the final year, after which it is clear from the relation between p, a, and i in (3) above that the value will have been reduced to zero [1] It is the rate of provision for depreciation, however, that reflects the cost of depreciation, and thus properly enters into Hicks's price equations. Hence, the duality of the price and quantity equations is preserved.

University of Kent at Canterbury

REFERENCES

1. HICKS, JOHN, *Capital and Growth*, Oxford, 1965.
2. KENNEDY, CHARLES, 'Time, interest and the production function', in J. N. Wolfe (ed.), *Value, Capital and Growth—Essays in Honour of Sir John Hicks*, Edinburgh, 1968.
3. MEADE, J. E., *A Neo-Classical Theory of Economic Growth*, George Allen and Unwin, 1961.
4. ROBINSON, JOAN, 'Depreciation', *Revista di Politica Economica*, Nov. 1959, reprinted in Joan Robinson, *Collected Economic Papers, Volume Two*, Basil Blackwell, Oxford, 1960
5. SRAFFA, PIERO, *Production of Commodities by Means of Commodities*, Cambridge, 1960

[1] For a fuller account of all this, see Meade [3], loc cit , Joan Robinson [4], and Sraffa [5], loc cit

THE CASE OF ADAM SMITH'S VALUE ANALYSIS[1]

By S. KAUSHIL

I

THIS paper attempts a reappraisal of Adam Smith's value analysis, which, according to many, would appear to show him up as a confused, and confusing, scholar, responsible for the almost endlessly varying and often contradictory meanings being put on his exposition on, and for the later, supposedly misdirected, progress of, the subject[2] A brief reference to the elements comprising the confusion is followed by a restatement of Smith's position, indicating the most likely culprit cause or causes, and the concluding remarks

The alleged confusion and/or inconsistency in Smith's analysis revolve around

(i) 'cause' versus 'measure' of exchange value,
(ii) labour-embodied versus labour-commanded measure of exchange value,
(iii) multiplicity of possible theories (causal explanations) of exchange value, viz. labour-embodied, labour-commanded, toil and trouble, three-factor cost of production, and demand-supply,
(iv) value theory in Smith's *Lectures* and that in his *Wealth of Nations*.

II

Adam Smith's conception of economic science, apart from being reflected in the over-all layout of the contents of the *Wealth of Nations* and the Plan of the Work, has been explicitly stated on p 643,[3] and, indeed, in the title

[1] Gratitude is due to Professor Vikas Mishra, but for whose Socratic obstetrics the paper would not have been, and to Professors Sir John Hicks, A K Das Gupta, R L Meek, and Joan Robinson for their learned criticism and useful suggestions, and for their encouragement

[2] E Roll, *A History of Economic Thought* (Faber edn), pp 156–60, esp. pp. 156–7, C Gide and C Rist, *A History of Economic Doctrines*, pp 92–6, also n 1 on p 94, J A Schumpeter, *History of Economic Analysis*, pp. 180, 307–11, J K Ingram, *A History of Political Economy*, pp 104–21, M Bowley, *Nassau Senior and Classical Economics*, pp 67–74, L. Robbins, *Robert Torrens and the Evolution of Classical Economics*, p 67, S Ambirajan, *Malthus and Classical Economics*, p 98, W A Weisskopf, *Psychology of Economics*, ch 4, esp pp. 38–43, 60–3, P H Douglas, 'Smith's theory of value and distribution', in *Adam Smith, 1776–1926*, esp. pp 77–102, E Kauder, 'Genesis of the marginal utility theory', *Economic Journal*, 1953, pp 638, 650, H M. Robertson and W L Taylor, 'Adam Smith's approach to the theory of value', *Economic Journal*, 1957, R A Macdonald, 'Ricardo's criticisms of Adam Smith', *Quarterly Journal of Economics*, 1912, pp 553, 556 The list is obviously incomplete, particularly in leaving unmentioned several other major histories of economic thought which carry and convey the same impression.

[3] All references to the *Wealth of Nations* are from the Modern Library reprint of Edwin Cannan's edition.

of the book: an inquiry into the nature and causes of the wealth of nations In labour, not, as with the Physiocrats, land, Smith finds the 'original' source of wealth,[1] and in division of labour, the cause of its enhancement. Division of labour, moreover, defines the nature of the economic system, emanating from the 'natural propensity to truck, barter, and exchange one good for another', it gives rise to the grand commodity exchange-complex, which is how the economic system has been and, indeed, can be conceived of. And the wealth of a nation is simply the totality of these commodity exchanges during the year. Indeed, his concern with the behaviour of the totality would appear to presuppose Smith's concern with the logic of the microcosmic commodity flows which lead up to the macrocosmic totality It is this logic which has been exposed in Book I of the *Wealth of Nations*, the chapters on division of labour spelling out the commodity-flow complex, the economic system, and the later chapters, on value and distribution, the logic proper. The exchange complex that characterizes the *economic system implies certain ratios of exchange or exchange values*—a social phenomenon—and the description of the rules and principles according to which these ratios or values are determined constitutes the subject-matter of Smith's value analysis [2]

Adam Smith begins, towards the end of chapter 4, Book I, by distinguishing, using the water-diamond illustration—and for the sole purpose of distinguishing[3]—between use-value and exchange value, and promises to present in the ensuing three chapters the analysis of exchange value in terms, respectively, of (i) the 'real' (invariable) measure of exchange value (ch 5), (ii) the causal components of exchange value (ch 6), and (iii) the deviations of the actual (market) price from the normal (natural) price (ch 7) Thus.

> In order to investigate the principles which regulate the exchangeable value of commodities, I shall endeavour to shew,
>
> First, what is the real measure of this exchangeable value, or, wherein consists the real price of all commodities.
>
> Secondly, what are the different parts of which this real price is composed or made up.
>
> And, lastly, what are the different circumstances which sometimes raise some or all of these different parts of price above, and sometimes sink them below their natural or ordinary rate, or, what are the causes which sometimes hinder the market price, that is, the actual price of commodities, from coinciding exactly with what may be called their natural price.[4]

[1] See below, p 65

[2] *Wealth*, p 28.

[3] That this illustration has been used solely to bring out the distinction between the two types of value, and not as some critics believe and assert (e g Schumpeter, op cit., p 309, Kauder, op. cit., p 650, Douglas, op cit., pp 78-81) to establish any relationship between the two, is self-evident Moreover, had this been Smith's intention here, he could have solved the so-called paradox, he had done so in his 1762 *Lectures* He simply wanted to sort out the subject-matter of his value analysis, viz exchange value

[4] *Wealth*, pp. 28–9.

Conceptually, (1) above is posterior to the emergence of the phenomenon of exchange value, and does not involve a 'theory', causal explanation, of exchange value, it rather involves the identification and quantification of existing exchange value For analytic purposes, however, measure, or rather the concept of measure, is a prerequisite for the explanation of exchange value, for unless one has a common measure of exchange value, one cannot quantify and make commensurate the exchange ratios between commodities, and unless one is able to do that, one cannot proceed with the analysis and causal explanation of the determination of these ratios. In any case, Smith does not mix up his metaphors and keeps the two issues, of measure and of cause, conceptually distinct although analytically integrated

Smith argues that the popularly comprehended and adopted measures, corn and silver, are not invariable in their own value and, therefore, inadequate and unsatisfactory for inter-temporal and inter-spatial comparisons of exchange value,[1] because

> as a measure of quantity, such as the natural foot, fathom, or handful, which is continually varying in its own quantity, can never be an accurate measure of the quantity of other things, so a commodity which is itself continually varying in its own value, can never be an accurate measure of the value of other commodities.[2]

He himself believes that, since the 'toil and trouble' undergone and the ease, comfort, and liberty sacrificed by an average worker during a certain labour time of average type, averaged for the differences in skills and hardships, remain of the same value to the worker irrespective of space and time, 'the quantity of labour which a commodity ought commonly to purchase, command or exchange for', expressing subjective disutility, toil, trouble, etc , is the real, invariable measure

> Equal quantities of labour, at all times and places, may be said to be of equal value *to the labourer*. In his ordinary state of health, strength and spirits, in the ordinary degree of his skill and dexterity, he must always lay down the same portion of his ease, his liberty, and his happiness. The price *which he pays* must always be the same, whatever may be the quantity of goods which he receives in return for it. Of these, indeed, it may sometimes purchase a greater and sometimes a smaller quantity; but it is their value which varies, not that of the labour which purchases them At all times and places that is dear which it is difficult to come at, or which it costs much labour to acquire; and that cheap which is to be had easily, or with very little labour. Labour alone, therefore, never varying in its own value is alone the ultimate and real standard by which the value of all commodities can at all times and places be estimated and compared. It is their real price, money is their nominal price only.[3] (Emphasis added.)

And again a little later,

> Labour, therefore, it appears evidently, is the only universal, as well as the only accurate measure of value, or the only standard by which we can compare the values of different commodities at all times and at all places.[4]

[1] *Wealth*, 31–3 [2] Ibid 32–3 [3] Ibid. 33. [4] Ibid. 36

Thus, if a commodity, or the totality of commodities, commands more of the labour time of average type than it did before or elsewhere, this commodity, or the totality, has, according to Smith, increased in value in real terms. Aware, however, of the need to have a more easily comprehensible objective counterpart of this labour-commanded measure, he suggests the corn wage-unit measure as being better than the money wage-unit measure, for, he holds, over distant times corn remains relatively more stable in its real (labour-commanded) value due to the relative constancy of subsistence corn wage, while money and silver do not, although he realizes that it would at best be an approximate counterpart. In his own words:

> Equal quantities of labour will at distant times be purchased more nearly with equal quantities of corn, the subsistence of the labourer, than with equal quantities of gold and silver, or perhaps of any other commodity. Equal quantities of corn, therefore, will, at distant times, be more nearly of the same real value or enable the possessor to purchase or command more nearly the same quantity of the labour of other people. They will do this, I say, more nearly than equal quantities of almost any other commodity; for even equal quantities of corn will not do it exactly.[1]

Smith does seem to have carved out a sufficiently valid device for measuring GNP and making its inter-temporal and inter-spatial comparisons, so imperative for his growth-centric essay.[2]

More important is, however, the fact that Smith never uses, not even by implication, the labour-commanded as a cause of value. It is always used as the measure. The concept of 'toil and trouble' has been used only to establish the validity of the labour-commanded measure as the universal invariable measure, and not to propound any causal explanation of exchange value.

Nor does he ever use his concept of labour-embodied as the measure of value. In the exceptional case of the one (labour)-factor model, where

> . . . the quantity of labour commonly employed in acquiring or producing any commodity, is the only circumstance which can regulate the quantity of labour which it ought commonly to purchase, command or exchange for . . .[3]

the cause (labour-embodied) and the measure (labour-commanded) do, in effect, appear to become identical, allowing for labour-embodied to be a measure as well as cause of value. This, indeed, is the rub, the source of

[1] *Wealth*, p. 35.

[2] Smith's treatment could, indeed, be interpreted as an attempt to answer the need for an index of economic progress. Having no statistics of GNP, and, more important, no index numbers, Smith would seem to have resorted to productivity per man-hour (PMH) as the index of economic progress. His intense concern with the value of silver and corn (in the second half of Book I, ch. 5, and in 'Digression' on the value of silver towards the end of Book I) would certainly suggest this. His labour-commanded measure would, on this interpretation, be just the reciprocal of PMH. M. Blaug has attempted to work out from Smith's treatment an index of economic welfare. See, M. Blaug, 'Welfare indices in the *Wealth of Nations*', *Southern Economic Journal*, Oct. 1959.

[3] *Wealth*, pp. 47–8.

all confusion on the point,[1] a confusion for which Smith is simply not to blame at all. For the two concepts, despite their interchangeability in the case of the one-factor model, are really distinct and separate. The interchangeability is, of course, due to labour-embodied being the only cause of value in the one-factor model. Thus, in the two-factor model, where capital has been accumulated, labour-embodied in a commodity is no longer 'the only circumstance that regulates' labour-commanded by it, capital (profits) being the other determinant,[2] while in the three-factor model, where land has been appropriated, land (rent) becomes the third determinant [3]

But even in the one-factor model, labour-embodied being the only cause is less than genuine, it is definitional, the boxes for the remaining two elements, capital and land, in Smith's conceptualization of the cause of value, being empty In the two (labour and capital)-factor model, too, labour-embodied can be used as both the measure and the cause of value. But this, again, is less than genuine, and is possible inasmuch as it is possible to convert the capital element into labour element, labour-embodied and labour-commanded once again becoming identical. It is only in the three-factor model, with land as one of the factors, that labour-commanded is visibly (and inescapably?) only the measure, and land, labour, and capital jointly the cause, of value. But, in Smith's conceptualization, labour-commanded has all along remained the only measure, and the three factors jointly the cause, of value irrespective of the dimension of the model Thus·

> In every society the price of every commodity finally resolves itself into some one or other, or all of those three parts, and in every improved society, all the three enter more or less, as component parts, into the price of the far greater part of commodities.[4]
>
> The real value of all the different component parts of price, it must be observed, is measured by the quantity of labour which they can, each of them, purchase or command. Labour measures the value not only of that part of price which resolves itself into labour, but of that which resolves itself into rent, and of that which resolves itself into profit.[5]

It may be pointed out that Smith's use of labour-commanded in 'toil and trouble' (or disutility) terms as the real, invariable measure of exchange value satisfies—intuitively, of course—the basic requirement for such a measure. that it should be a *datum* and extra-systemic to the causal system explaining the value phenomenon. The significance of this fact needs emphasizing, especially in view of its inadequate recognition, if not complete neglect, at the hands of Smith's critics and interpreters.

[1] It was on this point that Ricardo misunderstood, and misrepresented, Smith. See below, p 70.

[2] *Wealth*, 49. [3] Ibid. 49. [4] Ibid. 50 [5] Ibid. 50.

It may further be pointed out that while, as we shall see, Smith's conceptualization of the market price to natural price adjustment mechanism does involve the interdependence of factor market rates and product market prices, the natural rates of the three component factors are data to, i e are determined by parameters outside, the product natural price determination system, so that Smith's model could be said to be *determinate* [1]

Smith, beyond doubt, has a three-factor theory of exchange value Indeed, he explicitly develops a three-factor micro model of the exchange value of a single commodity measured in labour-commanded, and extends it to the three-factor macro model of the exchange value of the totality of commodities [2] He further extends it to denote his conception of the distribution schema in terms of the three factor shares as the 'three original sources of all revenue as well as of all exchangeable value', at both micro and macro levels [3]

Admittedly, Smith does emphasize the role of labour, and there are statements throughout Book I referring to labour as the 'original price' or 'purchase money' and as the source of the whole 'produce', which to the less careful reader might indicate a labour theory of exchange value A closer perusal shows that the statements which speak of labour as the 'original price' or 'purchase money' refer to the measure rather than the cause attribute of labour [4] In the other case, it is to be noted, labour is spoken of as the original source of '*produce*', not of *exchange value* For Smith, 'produce' always meant the physical, tangible output, this being the nature of the wealth of a nation. Initially, he took the position that labour was the original source of whole 'produce' But later, during the course of his analysis of the origin of rent, he seems to have come to regard land (nature) *also* to be an original source, along with labour, of 'produce' Thus, we find that in his chapter on rent (Book I, ch 11) and in all the later Books, he always speaks of the whole produce of land *and* labour And nowhere in the whole work does he ever talk of capital as an *original* source of produce But when he analyses the sources or components of exchange value of a commodity, he clearly formulates a three-factor explanation. His final position would thus emerge to be labour is not the only original source of 'produce', nor is it the only determinant of exchange value The claim by Wieser,[5] and those who follow him that

[1] Cf M Bowley, 'Some seventeenth-century contributions to the theory of value', *Economica*, 1963, p 136, where Smith's approach on the point is shown to be simply in tune with the going tradition Cf also, M Dobb, *Political Economy and Capitalism*, ch 1 on 'The requirements of a theory of value'.

[2] *Wealth*, pp. 50-2, 54

[3] Ibid 52-4.

[4] See, for example, *Wealth*, pp. 30, 33, see also Schumpeter, op cit., pp 309-11

[5] F von Wieser, *Natural Value*, Preface, pp xxvii–xxix, cf Schumpeter, op cit, p 189, last two sentences of n. 20

there are two theories, the philosophical labour theory and the empirical three-factor cost theory, of value coexisting in the *Wealth of Nations* would thus seem to be untenable.

Lastly, while there is no doubt a shift in Smith's treatment of value analysis in the *Wealth of Nations* compared with that in his *Lectures*, the shift, despite views to the contrary,[1] is for the better. In his *Lectures*, Smith had merely hinted at the relationship between natural price, which he then conceived of as natural wage price, and market price, and enumerated the three determinants of market price[2] The treatment in the *Wealth of Nations*, on the other hand, is fuller, more adequate and elegant, incorporating, and soaring far above, not only the bare-skeleton, pedagogic, treatment in the *Lectures* but also the utility-scarcity tradition long established by 1776[3]

Natural price is defined as the three-factor cost price which is to be paid if the commodity is to be produced and brought to the market. Thus:

> When the price of any commodity is neither more nor less than what is sufficient to pay the rent of the land, the wages of the labour, and the profits of the stock employed in raising, preparing, and bringing it to market, according to their natural rates, the commodity is then sold for what may be called its natural price.[4]

Market price is defined as the actual price paid in the market, and is determined by the interaction of 'effectual demand' with supply.

> The market price of every particular commodity is regulated by the proportion between the quantity which is actually brought to market, and the demand of those who are willing to pay the natural price of the commodity, or the whole value of the rent, labour, and profit, which must be paid in order to bring it thither. Such people may be called the effectual demanders, and their demand the effectual demand, since it may be sufficient to effectuate the bringing of the commodity to market.[5]

It (market price) may be equal to, or higher or lower than, natural price, depending upon the extent of 'effectual demand' relative to supply in the market. Thus, if there is excess demand, the extent to which market price is *above* natural price would depend upon. (i) the degree of 'greatness of the deficiency', i.e. the degree of scarcity, (ii) the relative 'riches and wanton luxury' of the buyers, and, if the buyers are equally rich, (iii) the relative eagerness for 'acquisition' of the commodity depending upon its

[1] Robertson and Taylor, op. cit., pp. 181–95, *passim*.

[2] Adam Smith, *Lectures on Justice, Police, Revenue and Arms* (ed E Cannan), section 7, esp. pp. 176–8.

[3] Those who find Smith's analysis inadequate and falling short of the already 'well-developed tradition' (e.g Schumpeter, op cit, pp. 308–9, Kauder, op. cit, pp. 638, 650, Douglas, op. cit, pp. 79–81, Robertson and Taylor, op. cit, pp. 181 ff) do him less than justice. They seem to over-appreciate the content of the 'tradition' and under-appreciate that of the *Wealth of Nations* In fact, none before Smith had as elaborately and explicitly brought out the operation of the market mechanism and the determination of market and natural prices and their relationships, as Smith does. He makes clear what his predecessors had been fumbling for.

[4] *Wealth*, pp 55–6.

[5] Ibid. 56

'importance' for the buyer, i e the intensity of demand for it depending upon its utility for the buyer. To quote:

When the quantity of any commodity which is brought to market falls short of the effectual demand, all those who are willing to pay the whole value of the rent, wages, and profit, which must be paid in order to bring it thither, cannot be supplied with the quantity which they want. Rather than want it altogether, some of them will be willing to give more. A competition will immediately begin among them, and the market price will rise more or less above the natural price, according as either the greatness of the deficiency, or the wealth and wanton luxury of the competitors, happen to animate more or less the eagerness of the competition. Among competitors of equal wealth and luxury the same deficiency will generally occasion a more or less eager competition, according as the acquisition of the commodity happens to be of more or less importance to them Hence the exorbitant price of the necessaries of life during the blockade of a town or in a famine.[1]

Smith's awareness of scarcity, income, and utility as the forces behind the effectual demand function is obvious. So is his awareness of the forces behind the supply function Thus, in the case of excess supply, the degree of 'greatness of the excess' and perishability, or durability, of the commodity, through the intensity of the sellers' competition, determine the extent to which the market price is *below* natural price Thus·

When the quantity brought to market exceeds the effectual demand . . . The market price will sink more or less below the natural price, according as the greatness of the excess increases more or less the competition of the sellers, or according as it happens to be more or less important to them to get immediately rid of the commodity. The same excess in the importation of perishable, will occasion a much greater competition than in that of durable commodities, in the importation of oranges, for example, than in that of old iron [2]

Despite the fact that he never intended to inquire exclusively into the nature and working of the market mechanism, his being a larger canvas and a much broader and grander objective, Smith's analysis of the process of adjustment of market price to natural price reveals his profound understanding of the functioning of the competitive market mechanism Cost determined natural price is conceived of as the 'centre of repose and continuance' to which the demand-supply determined market price 'continually' gravitates. In his own words:

The natural price, therefore, is, as it were, the central price, to which the prices of all commodities are continually gravitating. Different accidents may sometimes keep them suspended a good deal above it, and sometimes force them down even somewhat below it. But whatever may be the obstacles which hinder them from settling in this centre of repose and continuance, they are constantly tending towards it.[3]

This is the long-run stable equilibriuim, *à la* Marshall There is also a clear understanding, if only at a rudimentary level, of the interdependence of the commodity and factor markets Indeed, the process of adjustment is shown to be brought about via the necessary effects of any deviation of

[1] *Wealth*, p. 56. [2] Ibid. 57. [3] Ibid. 58.

market price from natural price on factor rewards and consequent adjustment of factor supply and product supply to demand. It would be worthwhile to quote Smith at length in this respect

> The quantity of every commodity brought to market naturally suits itself to the effectual demand It is the interest of all those who employ their land, labour, or stock, in bringing any commodity to market, that the quantity never should exceed the effectual demand, and it is the interest of all other people that it never should fall short of that demand.
>
> If at any time it exceeds the effectual demand, some of the component parts of its price must be paid below their natural rate. If it is rent, the interest of the landlords will immediately prompt them to withdraw a part of their land, and if it is wages or profit, the interest of the labourers in the one case, and of their employers in the other, will prompt them to withdraw a part of their labour or stock from this employment The quantity brought to market will soon be no more than sufficient to supply the effectual demand All the different parts of its price will rise to their natural rate, and the whole price to its natural price.
>
> If, on the contrary, the quantity brought to market should at any time fall short of the effectual demand, some of the component parts of its price must rise above their natural rate If it is rent, the interest of all other landlords will naturally prompt them to prepare more land for the raising of this commodity, if it is wages or profit, the interest of all other labourers and dealers will soon prompt them to employ more labour and stock in preparing and bringing it to market. The quantity brought thither will soon be sufficient to supply the effectual demand. All the different parts of its price will soon sink to their natural rate, and the whole price to its natural price [1]

All this is woven into an assumptional framework of self-interest *laissez-faire*, and competition.[2] Smith's perception, in the core, is, in this respect, almost Walrasian In fact, it emerges automatically from his conception of the economic system, and provides the basis for his support for *laissez-faire*, a basis which is altogether rational as against the Physiocratic ethico-emotional basis

Adam Smith's value analysis, thus, is not only free from the confusion and inconsistency of which it has been charged by the majority verdict but also has a grandeur of coverage Never before or after Adam Smith has value analysis in a single sweep performed so grand a feat That most interpreters proved myopic is an altogether different matter the most likely explanation for which is presented in the following section

III

We begin by taking note of those who presented a more faithful portrayal of the master's value analysis and thus attempted to redeem him of the charges. However, even they failed to appreciate its full relevance

[1] *Wealth*, pp 57–8, see also p. 62, second paragraph

[2] See *Wealth*, pp 59–62, where Smith shows how the market price may stay away from natural price, and the adjustment process checked, under conditions of monopoly, government regulations, imperfect knowledge and imperfect competition, thus implying that when these conditions are absent, i e perfect competition and *laissez-faire* prevail, the market mechanism would necessarily bring about the long-run equilibrium of 'repose and continuance'.

and import and, in particular, to explain what led to Smith's lead in this respect being lost sight of and, indeed, distorted all along.

There are only two such writers[1] and both point towards the same cause for Smith's analysis having been misread: the subjectivist bias originating in the 'marginal revolution' which caused the micro-ization of economic analysis and made the problem of valuation and resource allocation the core of economic science. The rationale of Smith's involvement with the invariable measure was not grasped by the marginalist interpreters so that his analysis of the measure, labour-commanded in toil and trouble terms, appeared to be brushing sometime against the three-factor cost theory of natural price and at other times against the misunderstood labour-embodied cause/measure of exchange value. The marginalists looked in Smith for a 'theory' of value where, according to Henderson, it was a subsidiary to the 'generic' and, according to Das Gupta, there was none for none had been intended.

Henderson and Das Gupta would appear partly to be right in suggesting that the reading of confusion between cause and measure and of inadequacy in Smith's 'theory' of value is to be attributed to the marginalist bias. But they would both appear to be inadequate in their own comprehension of the relevance, significance, and adequacy of the theory of value proper as developed by Smith. Henderson constructs no more than an outlinish picture of Smith's value analysis while Das Gupta presents what is at best a profile. Henderson does not seem to see fully the profundity and versatility of the master's performance. Das Gupta, on the other hand, takes the view that value analysis in the *Wealth of Nations*, an essay in growth economics, could have no purpose other than to evolve an invariable measure for inter-temporal and inter-spatial comparisons of national product, a purpose Smith succeeds in achieving. It does seem that in his anxiety to straighten the rod Das Gupta has bent it a little too much to the other side.

The reading of confusion and inconsistency regarding labour-embodied/labour-commanded measures of exchange value could not, however, have been a product of the marginalist bias. Actually, we find David Ricardo accusing Smith of inconsistency in this respect almost half a century prior to the 'marginal revolution'. He chided Smith for having inconsistently and unnecessarily shifted away from the labour-embodied as 'the only rule' in the 'early and rude state', i.e. the one-factor model, to the labour-commanded measure in the advanced state, i.e. the multi-factor model.[2]

[1] J. P. Henderson, 'The macro and micro aspects of the Wealth of Nations', *Southern Economic Journal*, 1954, pp. 25–35; A. K. Das Gupta, 'Adam Smith on Value', *Indian Economic Review*, 1960–1, pp. 105–15.

[2] P. Sraffa and M. Dobb (eds.), *Works and Correspondence of D. Ricardo*, vol. 1, pp. 12 et seq.

This, indeed, was the source of all subsequent misunderstanding, and failing to comprehend Smith's position Ricardo misled himself into cor mitting the original sin of misunderstanding and misrepresenting Smit He did not consider at all Smith's 'toil and trouble' idea and went ahea with his critique of the latter's measure approach, as he saw it, and t establish his own labour-embodied concept as the 'nearest approximatio to an invariable measure. His eagerness to establish his own point wa perhaps also responsible for Ricardo's having come to present a wron misconstrued interpretation of Smith's value analysis so far as the latter measure approach is concerned [1]

Admittedly, Smith's value analysis had been criticized even earlier, e g by Say and Lauderdale, but that had been at a different plane altogeth and there had been no charge of confusion and inconsistency Aft Ricardo and before the 'marginal revolution' there had been a lot o writers, both disciples and critics of Ricardo, including Marx, who rea similar inconsistency in Smith's analysis. Apparently, they had viewed through Ricardian glasses, for, eventually, Ricardo had come to hav profound intellectual sway over the economic fraternity, and had come t be believed, he himself having contended almost as much, to have lai down the basic principles of economic analysis on the basis of a correctiv and more logical reformulation of the *Wealth of Nations.*[2] This must no however, be taken to imply that Ricardo's reformulation of Smithia principles was uncalled for, or erroneous; only that he misrepresente Smith in the process

And the meaning and rationale of Ricardo's own approach to measur and his analysis of the invariable measure were only too often confounde with a 'theory' of value The effect was twofold (i) What Ricardo sai to be inconsistency in Smith's analysis of measure came to be accepted a such without anyone detecting his misconceptions on the point (ii) Whe holes came to be detected in Ricardo's analysis itself, these were regarde as scarcely more than amplified manifestations of what originally wer Adam's sins [3] What the marginalist bias did was to compound thi already prevalent attitude towards Smith's value analysis by adding to i a new dimension So, when judgement came, Smith had to suffer both fo Ricardo's misrepresentation and for the neo-classical bias, and the verdic of confusion and inconsistency and inadequacy went on piling up

When economic analysis came to be macro-ized under the impact of th 'Keynesian revolution', in contrast to the neo-classical micro-ization, i was the growth aspect of Smith's analysis that came to be the focus o

[1] Cf. MacDonald, op cit, pp 585–6

[2] See, for example, Schumpeter, op. cit., pp 472, 474, 482

[3] For instance, see Ingram, op cit, pp. 104–5, 120–1, Schumpeter, op. cit, p. 310 Bowley, op cit, p 73, Robertson and Taylor, op cit, p 189

attention of the interpreters, and the import of his value analysis came tc be recognized in the growth-centric context alone, as a provider oi 'measure', the other, 'theory', facet not getting its due. Much can yet b(learnt from the father of economic science about an integrative approacl combining the analysis of the macrocosmic problems of economic growtl and stability, on the one hand, and the microcosmic problems of valuatioi and resource allocation, on the other

Kurukshetra University, India

A MODEL OF INTERSECTORAL MIGRATION AND GROWTH

By ANDREU MAS-COLELL *and* ASSAF RAZIN

I. Introduction

OUR main purpose in this paper is to show how some of the patterns of growth of a dual economy studied by Fei and Ranis [4], Jorgenson [6, 7], and Dixit [2] can be explained by a simple neoclassical growth model Such patterns of growth include: a decreasing rate of migration from rural to urban sector; a stage of accelerated accumulation of capital, etc In order to accomplish this objective it is necessary to introduce explicitly migration into the framework of neoclassical growth models Therefore we shall assume that labour cannot be transferred instantaneously between sectors Furthermore, the rate of migration will be determined by economic forces In this respect our model is different from the one studied by Harris and Todaro [5] in which the industrial wage is institutionally fixed. There is an unavoidable trade-off between simplicity and generality. Since we regard the simplicity of the model as one of its most compelling aspects we shall not attempt to generalize and we shall use freely special assumptions For example, following Jorgenson [6] we shall assume that production functions are of Cobb–Douglas form [1]

The plan of this paper is as follows In Section II we shall introduce a model of migration and capital accumulation for an economy with agricultural and industrial sectors. We shall show that the direction of migration between sectors is completely determined by the proportion of the total labour force occupied in agriculture In Section III we shall analyse rates of growth of migration, capital accumulation, industrial output–capital ratio, and the terms of trade In Section IV we shall discuss a policy of subsidy-tax. The steady states generated by various tax rates will be derived and those which are 'inefficient' will be singled out. We shall also analyse the effect of policy on the length of the period of migration

II. The model

Consider an economy with two productive sectors, an agricultural sector A producing output for consumption and an industrial sector I producing output for consumption and investment.[2] The production functions are

[1] A more general model with wage differential was studied in a different context by Bosch and the authors [1]

[2] The reader may be referred to Uzawa [10] who studied a neoclassical two sector model of economic growth.

assumed linear homogeneous of Cobb–Douglas form. The total labour force is a constant fraction of total population (assumed for convenience to be one). Thus, *per capita* outputs, denoted by y, can be written [1]

$$y_I = \rho k_I^{\beta}, \qquad y_A = (1-\rho)k_A^{\alpha}, \tag{1}$$

where k_I, k_A are capital–labour ratios in sector I and A respectively and ρ is the proportion of total labour force employed in the industrial sector.

There exists full employment of capital and labour,

$$\rho k_I + (1-\rho)k_A = k, \tag{2}$$

where k is the capital–labour ratio available for the economy Let p be the price of the industrial good in terms of agricultural good. Assuming that capital is instantaneously transferred, competition will equalize the marginal productivity of capital in both sectors

$$p\beta k_I^{\beta-1} = \alpha k_A^{\alpha-1} = r. \tag{3}$$

Denoting the wage rate by w, competition within sectors implies

$$w_I = p(1-\beta)k_I^{\beta}, \qquad w_A = (1-\alpha)k_A^{\alpha} \tag{4}$$

Per capita national income (in units of agricultural good) is expressed by y_A+py_I Postponing a more extensive discussion for the next paragraph, we shall now assume, for concreteness, a constant ratio of saving to income s, and also that a constant proportion of income δ is being spent on the industrial good for consumption purposes Therefore, demand for industrial output is $s(y_A+py_I)+\delta(y_A+py_I)$, supply of industrial output is py_I. In equilibrium we have

$$(\delta+s)(y_A+py_I) = py_I \tag{5}$$

Let $\lambda = s/s+\delta$ be the proportion of total industrial output in the form of new capital goods. The solution of (1)–(3) and (5) is given by

$$k_I = \theta\frac{k}{\rho}, \qquad k_A = (1-\theta)\frac{k}{1-\rho} \tag{6}$$

where $$\theta = \frac{\beta s}{\beta s+\alpha(\lambda-s)} = \frac{(s+\delta)\beta}{(s+\delta)\beta+(1-s-\delta)\alpha}.$$

This shows that the momentary equilibrium is well defined

The model accommodates a more general situation where the fraction of the aggregate demand for industrial goods in total income is not necessarily constant over time Assume that the population consists of three groups of consumers the wage earners in the agricultural sector A, the wage earners in the industrial sector I, and the owners of capital C. Every group allocates a constant proportion of its income (not necessarily the same between groups) to consumption and investment in industrial goods

[1] The multiplicative constants of the Cobb–Douglas functions are eliminated by an appropriate choice of the units of labour and agricultural output

Let ν_i, μ_i stand for the proportion of income of group i spent on industrial goods for consumption and investment purposes, respectively; define $\chi_i = \nu_i + \mu_i$, $i = A, I, C$. Then, for example, if $\nu_I > \nu_A$, i.e. if the propensity to consume industrial goods is greater for industrial workers as compared with agricultural workers, the proportion of aggregate income spent on industrial goods for consumption purposes will increase when migration takes place.

Therefore, *per capita* demand for industrial output is

$$(1-\rho)\chi_A w_A + \rho\chi_I w_I + \chi_C rk;$$

supply of industrial output is py_I. In equilibrium we have

$$(1-\rho)\chi_A w_A + \rho\chi_I w_I + \chi_C rk = py_I. \tag{7}$$

Substituting (1), (2), (3), and (4) into (7) we obtain the same expression as in (6), where θ is substituted by

$$\theta' = \frac{(1-\alpha)\beta\chi_A + \alpha\beta\chi_C}{(1-\alpha)\beta\chi_A + \alpha(1-\chi_I(1-\beta))}. \tag{8}$$

Similarly, easy computation shows that the proportion of total industrial output in the form of new capital goods is a constant λ' given by

$$\lambda' = \frac{(1-\theta')\beta(1-\alpha)\mu_A + \theta'(1-\beta)\alpha\mu_I + \beta\alpha\mu_C}{\alpha\beta'}. \tag{9}$$

Hence the simple and the more general models are formally similar. Throughout the rest of the paper we shall analyse the simple model.

Population is increasing at a constant relative rate n. Therefore, capital–labour ratio will be accumulated according to

$$\dot{k}/k = \lambda y_I/k - n \tag{10}$$

$$= \lambda\theta^{\beta}\left(\frac{k}{\rho}\right)^{\beta-1} - n$$

where use has been made of the momentary equilibrium conditions in (6).

We assume that migration of labour[1] is positively related to wage differential

$$\dot{\rho}/\rho = f(w_I, w_A), \tag{11}$$

where f is a continuously differentiable function such that

$$\operatorname{sign} f = \operatorname{sign}(w_I - w_A).$$

The motion of the system can be conveniently analysed in the phase diagram of Fig. 1.

The stationary law of migration is solved by setting the right-hand side

[1] If M is the rate of migration into the industrial sector and L_I is the industrial labour then

$$\therefore \quad \frac{M}{L_I} = \frac{\dot{L}_I - nL_I}{L_I} = \frac{\dot{L}_I}{L_I} - n = \frac{\dot{\rho}}{\rho}.$$

of (11) equal to zero, i.e $w_I = w_A$. Using (4) and (6) we conclude that no migration will take place if·

$$\hat{\rho} = \frac{\alpha(1-\beta)\theta}{\theta(1-\beta)\alpha+(1-\theta)(1-\alpha)\beta}.^{1} \tag{12}$$

From (11) we see that $\quad \rho \gtrless 0 \quad$ as $\rho \lessgtr \hat{\rho}$. (13)

We illustrate this finding with a numerical example. Let $s = 0{\cdot}15$, $\lambda = 0{\cdot}20$, $\delta = 0{\cdot}60$, $\alpha = 0{\cdot}30$, and $\beta = 0{\cdot}40$, then $\hat{\rho} \approx 0{\cdot}70$ Therefore as

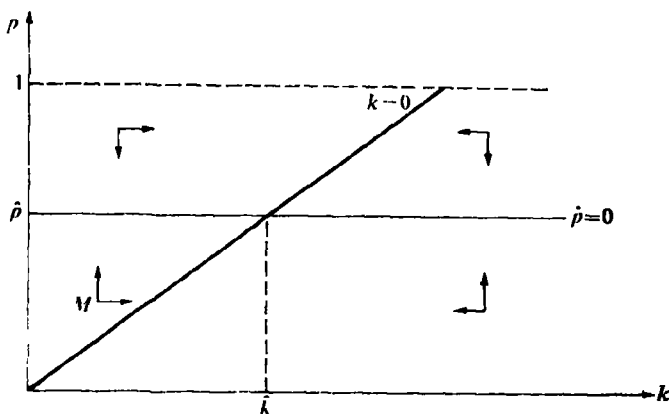



Fig 1

long as the proportion of population occupied in industry is less than 0·70 migration to the industrial sector will take place [2]

The differential equation (10) has a stationary solution when

$$k = c\rho, \tag{14}$$

where

$$c = \left(\frac{\lambda}{n}\right)^{1/(1-\beta)} \theta^{\beta/(1-\beta)}.$$

As can be seen from Fig 1 and (10) and (13) the economy has a unique and globally stable steady state $\rho = \hat{\rho}$, $\hat{k} = c\hat{\rho}$.

III. Relative rates of growth

In order to pursue the analysis further we need a specific form for the migration equation (11) The following migration equation is similar to the ones used by M. Todaro [9], P Zarembka [11], and satisfies the plausible requirement that $\rho/\rho \to \infty$ as $w_I/w_A \to \infty$ Moreover, this equation behaves nicely in our model.

$$\rho/\rho = \gamma\left[\frac{w_I - w_A}{w_A}\right] \tag{15}$$

where γ is a positive constant.

[1] In general, when production functions are not of Cobb–Douglas form the proportion $\hat{\rho}$ will depend on k See [1]

[2] In the more general model let $\alpha = 0{\cdot}3$, $\beta = 0{\cdot}4$ $\mu_1 = \mu_I = 0$, $\nu_A = 0{\cdot}4$, $\nu_I = 0{\cdot}75$, $\mu_C = 0{\cdot}1$, then, again, $\rho \approx 0{\cdot}7$

Hereafter we shall restrict our analysis to region M in Fig. 1, where capital is being accumulated and migration of labour into industry takes place. We shall discuss in this section the implications of the model for relative rates of growth of migration, capital accumulation, output–capital ratio in the industrial sector, and the terms of trade.

1. The relative rate of growth of migration ($\dot{\rho}/\rho$).[1] Substituting (3)–(4) and (6) into (15) we obtain

$$\dot{\rho}/\rho = \gamma\left[\frac{(1-\beta)}{\beta}\frac{\alpha}{(1-\alpha)}\frac{(1-\rho)}{\rho}-1\right]. \tag{16}$$

Clearly the rate of growth of migration decreases when migration into the industrial sector takes place (i.e. ρ increases).

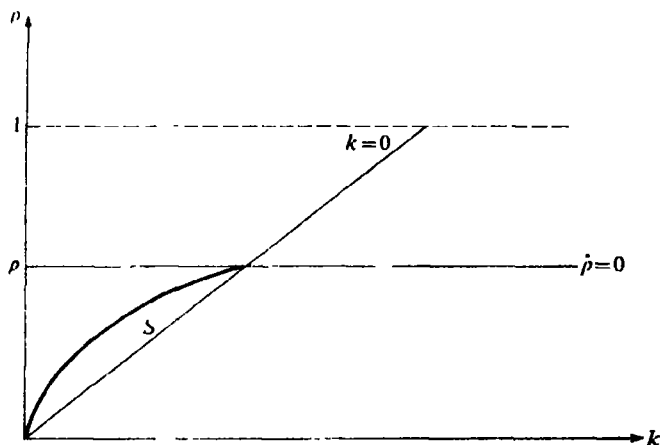



Fig. 2

2. The relative rate of growth of capital accumulation. Denoting this rate by $\dot{K}/K$ we have $\dot{K}/K = \dot{k}/k+n$. Differentiating (10) with respect to time[2] and taking (16) into account we can draw in Fig. 2 the locus

$$\frac{d(\dot{k}/k)}{dt} = 0.$$

This locus cannot intersect the $\dot{k} = 0$ or $\dot{\rho} = 0$ curves. (Furthermore, for any k there exists a ρ such that (k, ρ) is below this curve and for any ρ there exists a k such that (k, ρ) is to the left of it.)

Referring to Fig. 2 any path which starts initially in region S will exhibit

[1] Observe that $\dot{L}_I/L_I$, the rate of growth of industrial labour force, is equal to $(\dot{\rho}/\rho)+n$. Recently Sato and Niho [8] have studied a model of a dual economy where population growth is related to the level of *per capita* income. Their expression for rate of migration (equation (17)) therefore includes as an argument also *per capita* income. Their approach, however, poses the problem that changing the numeraire by which *per capita* income is measured will change the pattern of migration they attempt to study.

[2] From (10) we have $\frac{d\dot{k}/k}{dt} = (\dot{k}/k)\lambda(1-\beta)(\dot{\rho}/\rho-\dot{k}/k)$

initially a phase of accelerated capital accumulation but eventually will enter a phase of decreasing rate of growth of capital as the economy approaches the steady state. This is shown in Fig. 3.

3 The industrial output–capital ratio (z). From (1) and (6)

$$z = \theta^{\beta-1}(k/\rho)^{\beta-1}.$$

Comparing this expression with (10) yields the conclusion that $\dot{K}/K$ and z move together over time as it is indicated in Fig. 3

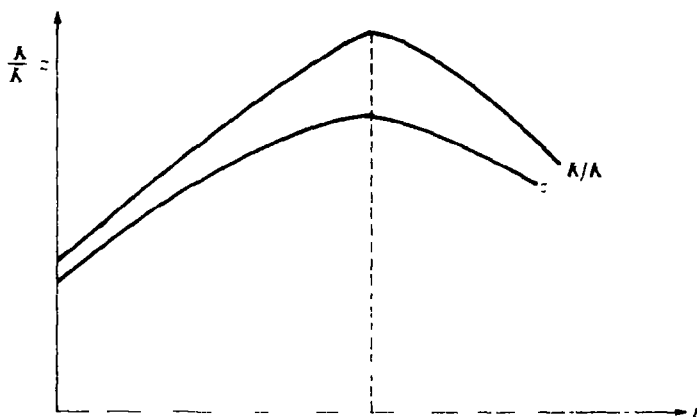



Fig. 3

4. The terms of trade of agriculture ($1/p$) From (3) and (6) we see that the terms of trade of agriculture are proportional to $k^{\beta-\alpha}\rho^{1-\beta}(1-\rho)^{\alpha-1}$. Clearly if the industrial sector is more capital intensive then the terms of trade will move monotonically in favour of agriculture

In particular our model is capable of generating the patterns of growth of migration and capital accumulation reported by Dixit [2]

IV. Policy implications

1 We shall show briefly how policy variables may be introduced in a simple way into the model Suppose an *ad valorem* subsidy (tax) at the rate of τ is given to the agricultural sector. Suppose that the government raises (gives) these funds from an income tax

Equation (3) becomes

$$(1+\tau)\alpha k_A^{\alpha-1} = p\beta k_I^{\beta-1}, \quad \tau > -1 \tag{17}$$

The rest of the model is unchanged

Let us define $$\theta_\tau = \frac{s\beta}{s\beta+(\lambda-s)\alpha(1+\tau)}.$$

In equation (6) we substitute θ_T for θ

In Fig. 4 we represent the locus of possible steady states as the subsidy rate τ ranges over $(-1, +\infty)$.[1]

[1] This is equivalent with θ_T ranging over (0, 1)

Point A in Fig. 4 represents a steady state corresponding to a higher subsidy rate than the steady state represented by B. However, some points on this locus will represent steady states which are inefficient for the economy. We say that a steady state is efficient if there are no other steady states with more of consumption (*per capita*) of both goods [1]

Per capita consumption of industrial goods increases monotonically with ρ along this locus (as can be inferred from (1) and (6)). *Per capita* consumption of the agricultural good is maximized when $\rho = \alpha$ on this locus.[2] Therefore any steady state where $\rho < \alpha$ is inefficient.

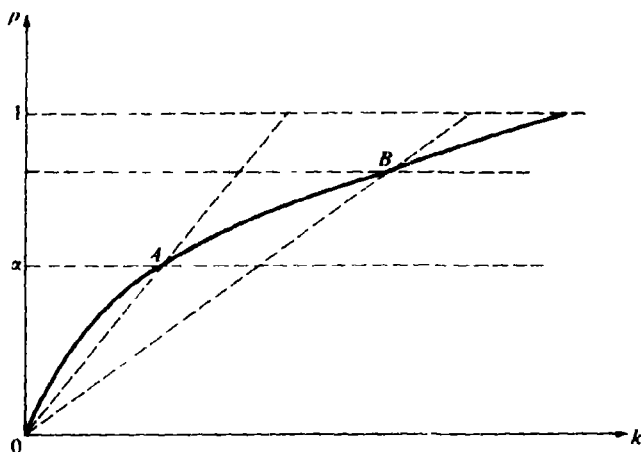



Fig. 4

2. The simplicity of the model enables us to solve explicitly for the time paths of ρ and k. This means that the period of time needed to reach some objectives of ρ and k, which can be controlled by the government through τ, is readily computable The solution for the differential equation (16)[3] is given by

$$\rho_t = (\rho_0 - \hat{\rho}) e^{-\gamma(\hat{\rho}/1-\hat{\rho})t} + \hat{\rho} \tag{18}$$

The explicit time path of k_t can be easily found by using the transformation $q = k^{\beta-1}$.

Suppose an amount of time $\bar{t}$ is needed for the economy to reach a given composition of labour force $\bar{\rho}$. What will be the effect of an increase in the rate of subsidy on the period in which the same labour force composition will be reached? To answer this question we differentiate totally the

[1] If, as in the model studied by Dixit [3], saving by market forces is assumed to be socially suboptimal then points on the locus of possible steady states other than the efficient points may become targets for the project evaluator

[2] This is the well-known 'Golden Rule'. Among steady states, $\lambda \rho k_I^{\beta} = nk$ the consumption of agricultural good is maximized when $\beta \lambda k_I^{\beta-1} = n$ Combining these with (2)–(4) we get $\rho = \alpha$.

[3] Observe that the differential equation (12) does not depend on k

right-hand side of (18) and set it equal to zero, thereby obtaining $dt/d\tau > 0$. Thus, subsidizing agriculture will lengthen the migration period

The University of Minnesota and Tel Aviv University

REFERENCES

1. BOSCH, A., MAS-COLELL, A., and RAZIN, A., 'Instantaneous and non-instantaneous adjustment to equilibrium in two-sector growth models', *Metroeconomica* (forthcoming).
2. DIXIT, A., 'Growth patterns in a dual economy', *Oxford Economic Papers*, 22 (2) July 1970, pp. 229–33.
3. —— 'Short-run equilibrium and shadow prices in the dual economy', *Oxford Economic Papers*, 23 (3), Nov. 1971, pp. 384–400.
4. FEI, J. C. H., and RANIS, G., *Development of the Labor Surplus Economy* Homewood· Irwin, 1964.
5. HARRIS, J. R., and TODARO, M. P., 'Migration, unemployment and development a two-sector analysis', *The American Economic Review*, 60 (1), Mar 1970, pp. 126–42.
6. JORGENSON, D. W., 'Testing alternative theories of the development of the dual economy', in Adelman and Thornbecke (eds.), *The Theory and Design of Economic Development*, Baltimore. Johns Hopkins Press, 1966, pp. 45–60.
7. —— 'Surplus agricultural labour and the development of a dual economy', *Oxford Economic Papers*, 19 (3), Nov. 1967, pp. 288–312.
8. SATO, R., and NIHO, Y., 'Population growth and the development of a dual economy', *Oxford Economic Papers*, 23 (3), Nov. 1971, pp 418–36.
9. TODARO, M. P., 'A model of labor migration and urban unemployment in less developed countries', *The American Economic Review*, 59 (1), Mar 1969, pp. 138–48.
10. UZAWA, H., 'On a two-sector model of economic growth', *The Review of Economic Studies*, 29 (1), 1961, pp. 40–7.
11. ZAREMBKA, P., 'Labor migration and urban unemployment: comment', *The American Economic Review*, 60 (1), Mar. 1970, pp. 184–6.

ESTIMATING THE IMPACT OF TARIFF MANIPULATION: THE EXCESS DEMAND AND SUPPLY APPROACH

By RICHARD BLACKHURST[1]

THE subject of this paper is a presentation of an approach to the problem of predicting the impact of tariff manipulation on both the country which is changing its tariffs and on the other countries involved, either as buyers or sellers in the world market for the commodities whose tariffs are being changed [2,3] The paper is concerned with the problem of quantifying the impact of tariff changes on the level of imports, and does not deal with the question of the welfare effects of tariff manipulation Because of its generalized nature the model is applicable to both non-discriminatory and discriminatory tariff changes, and, with minor modification, it is applicable to situations in which the relation between domestic and foreign prices is disturbed by means other than tariff manipulation [4]

The basic model

Let the world economy be divided into three groups· A, an aggregate of the importing countries which are simultaneously changing their tariffs on the commodity, B, an aggregate of the exporting countries to which the tariff change is applicable, and C, a grouping of all other countries [5]

There are three principal steps in the development of the basic model· (1) the derivation of an expression for estimating the price elasticity of A's demand for imports of the commodity, (2) the derivation of an expression for estimating the price elasticity of B's supply of exports of the

[1] Much of the work on the refinement of the model presented in this paper was carried out during the academic year 1968 9 when I held the position of Scholar in Residence at the United States Tariff Commission It should not be construed that the Tariff Commission necessarily concurs with the contents or conclusions of the paper I would like to thank Harry G Johnson and Arnold Harberger for their comments and suggestions on earlier drafts

[2] The original stimulus to my work on this model was provided by the work of Harry G Johnson [**10**, pp 46–74] and Arnold Harberger [**9**]. This paper draws on Part II of [**4**]

[3] The excess demand and supply approach has a long history in the literature, but its development has remained on a relatively simple level See, in addition to the works cited above, [**1**], [**2**], [**3**], [**7**], [**8**], [**18**], [**19**], [**22**], and [**24**] A model related to the one presented in this paper has been developed independently by Kiyoski Kojima [**13**]

[4] The model does not allow for balance of payments or exchange rate considerations In addition, it is not directly applicable to instances in which a country enters the world market for a commodity for the first time as a result of having received/granted a tariff concession

[5] The C grouping therefore includes importing countries which are not changing their tariffs, exporting countries to whom the tariff change is not applicable, and countries which neither import nor export the commodity The lower-case form of each letter is used throughout the paper as a subscript to indicate the aggregate or group to which a variable refers.

commodity, and (3) the derivation of an expression in which the two elasticities and the tariff change may be combined to obtain an estimate of the change in the value of A's imports.

The elasticity of the (excess) demand for imports

For any commodity imported into A we can write

$$D_a = Q_a + M_b + M_c, \tag{1}$$

where D and Q represent domestic consumption and domestic production respectively, and M represents imports.[1] Rearranging and converting to elasticity form we obtain

$$\eta_m = \frac{D_a}{M_b}\eta_a - \frac{Q_a}{M_b}e_a - \frac{M_c}{M_b}\epsilon_c \tag{2}$$

in which η_m is the elasticity of A's demand for imports from B, η_a and e_a are A's domestic demand and domestic supply elasticities respectively, and ϵ_c is the elasticity of (excess) supply of imports of the commodity from C (that is, from the exporting countries in the C grouping).[2]

The traditional expression for η_m has only the first two terms of equation (2) The three-term version is preferable, not only because it is more general, but also because it frequently is more appropriate to actual tariff changes This is true, for example, with preferential tariff changes, with most-favoured-nation reductions when certain countries are not eligible for MFN treatment, and in the case of tariff reductions when imports from certain countries already enter duty free under an existing preferential agreement.[3]

[1] It is assumed that each commodity is homogeneous in the sense that buyers do not discriminate on the basis of country of origin or brand names See [4] for a brief discussion of why this assumption may not be as restrictive as is generally assumed

[2] If equation (1) is rearranged, and P is used to represent price, we may write

$$\frac{\Delta M_b}{\Delta P} = \frac{\Delta D_a}{\Delta P} - \frac{\Delta Q_a}{\Delta P} - \frac{\Delta M_c}{\Delta P}$$

and

$$\left(\frac{P}{M_b}\cdot\frac{\Delta M_b}{\Delta P}\right) = \left(\frac{\Delta D_a}{\Delta P}\cdot\frac{P}{D_a}\right)\frac{D_a}{M_b} - \left(\frac{\Delta Q_a}{\Delta P}\cdot\frac{P}{Q_a}\right)\frac{Q_a}{M_b} - \left(\frac{\Delta M_c}{\Delta P}\cdot\frac{P}{M_c}\right)\frac{M_c}{M_b}$$

from which equation (2) follows directly The three elasticities on the right-hand side of equation (2) are weighted averages, as are the three corresponding elasticities in equation (4) The demand elasticities are negative values

[3] It is sometimes argued that equation (2) is not very helpful because it simply replaces a guess about the value of η_m with guesses about the values of the two domestic elasticities (the criticism has been directed to the two-term version since that is the version commonly encountered) See, for example, [**15**, pp 194–5] This criticism is inappropriate for two related reasons In many instances it is possible to use econometric estimates of the domestic elasticities rather than 'guessing' their values (The option of using econometric estimates of η_m—thus avoiding equation (2) altogether—suffers from the fact that such estimates generally are available only for highly aggregated commodity groupings, as well as from alleged weaknesses in the estimating techniques; on the latter point see [**9**], [**19**], and [**15**, esp pp 28–35]) Second, if estimates of the domestic elasticities are not available, most researchers would agree that there is a relatively narrow range of plausible values of the

The elasticity of the (excess) supply of exports

We may write for any commodity exported by B

$$Q_b = D_b + X_a + X_c, \tag{3}$$

where X represents exports. Rearranged and converted to elasticity form, this becomes

$$\epsilon_x = \frac{Q_b}{X_a} e_b - \frac{D_b}{X_a} \eta_b - \frac{X_c}{X_a} \eta_c \tag{4}$$

where ϵ_x is the elasticity of B's supply of exports to A, e_b and η_b are B's elasticities of domestic supply and domestic demand respectively, and η_c is the elasticity of (excess) demand for B's exports of those importing countries which are not altering their tariffs on the commodity.[1]

As with equation (2), the three-term version of equation (4) is more general and frequently more relevant than the version which expresses the export supply elasticity as a function only of the first two terms. For example, in any particular tariff changing round there are likely to be importing countries which do not participate, importing countries which exempt certain products, and—in the case of tariff reduction—importing countries which already allow certain products duty free entry.

The impact of the tariff change on the value of imports

Let P and Q represent price and quantity respectively, and t represent the weighted average *ad valorem* tariff. We may then write the following import demand and export supply equations

$$\log Q = \alpha + \eta_m \log(P\tau),$$
$$\log Q = \beta + \epsilon_x \log P,$$

where $\tau = (1+t)$. Since

$$d\log Q = \epsilon_x d\log P = \eta_m(d\log P + d\log\tau)$$

then

$$d\log P = \frac{\eta_m(d\log\tau)}{\epsilon_x - \eta_m}, \tag{5}$$

or

$$d\log P + d\log Q = \frac{\eta_m(\epsilon_x+1)}{\epsilon_x - \eta_m} d\log\tau$$

and

$$d\log V = \frac{\eta_m(\epsilon_x+1)}{\epsilon_x - \eta_m} d\log(1+t) \tag{6}$$

where V is the money value of imports of the commodity from B.[2] Not

domestic elasticities, conversely, it is not possible to define a similar narrow range c plausible values of an import demand elasticity (unless, of course, one has an implici version of equation (2) in mind)

[1] The derivation parallels the derivation of equation (2).

[2] The *net* change in A's imports of the commodity equals the change in imports from] plus the change in imports from C. Because the estimates for each individual commodit flow from a partial equilibrium model, they can be aggregated to obtain an estimate of th total impact of the changes in several tariffs only on the assumption that cross elasticitie in demand and supply may be neglected.

that equation (5) is an expression for the change in the price B receives for her exports to A.

The change in imports from B may be broken down into the consumption, production, and trade diversion effects in A and B The three effects in A are identified on the basis of the relative importance of each term in equation (2) to the value of η_m, while in B the three effects are determined by the relative importance of each term in equation (4) to the value of ϵ_x.[1]

Lacking an expression such as equation (6), researchers have in practice followed one of three alternatives—the principal attraction in each instance being the ease with which the change in imports can be estimated The first, and most common, assumes that imports are available at a constant (export) price (i e that $\epsilon_x = \infty$) [2] The second assumes that the change in imports will not affect A's domestic market price (i.e that $\eta_m = -\infty$).[3] The third alternative represents a small step in the direction of equation (6); in this instance, a particular pre-determined proportion of the tariff change is assumed to be passed on to B in the form of a higher or lower price (depending on whether the tariff is being lowered or raised), the remainder being passed on to A's consumers [4,5]

Equation (6) draws its significance from the fact that it removes the need to make highly simplified and/or arbitrary *a priori* judgements about the elasticity of one of the two curves This is important for the obvious reason that in many tariff-changing rounds the commodities will lie along a continuous spectrum of pairs of values of η_m and ϵ_x, the values varying according to which commodity, countries, and tariff-changing strategy are involved.[6]

When equation (6) is applied to the discrete changes encountered in actual tariff manipulation, it is necessary to allow for the fact that the two

[1] The three effects can be measured unambiguously only in physical units This follows from the fact that there are four possible prices to use in valuing the three effects (1) the original world price, (2) the original tariff-inclusive price, (3) the new price paid to B, and, with respect to trade diversion, (4), the new price paid to C exporters If the change in V obtained from equation (6) is divided on the basis of the relative contribution of each term to equations (2) and (4), the three effects in each area are valued at the new price paid to B.

[2] See, for example, [20] and [22]

[3] I was not able to locate an example of an empirical study in which the estimates were based on this assumption It is, however, mentioned often in general discussions—for example in connection with the issue of tariff preferences for the exports of the less-developed countries

[4] In [1], [2], and [3] Balassa and Kreinin use both the first and third alternatives, in the case of the latter, the *a priori* assumption about the effect of the tariff change on the price paid for imports is based in part on Kreinin's analysis of the impact of tariff concessions granted by the United States in 1955 and 1956—his conclusion being that 'It appears plausible that close to half of the benefit from tariff concessions granted by the United States accrued to foreign exporters in the form of increased export prices' [14, p 317]

[5] The estimated percentage change in the value of imports is obtained by multiplying the percentage change in one plus the tariff by η_m (when $\epsilon_x = \infty$) or $\epsilon_x - 1$ (when $\eta_m = -\infty$), the third alternative is nearly as easy, requiring only two additional simple steps See [4, Appendix II]

[6] See [4] for a discussion of the application of equation (6) to situations in which tariff reductions are coupled with the imposition of quotas

excess elasticities will not remain constant. That is, while it is not unre
able to assume that the three underlying elasticities in equation (2)
equation (4) remain constant *over the range of the change*, it is clear tha
values of the three ratios in each equation will be changing continu
as the tariff change affects the level of D, Q, M, and X.[1] One soluti
to assume the tariff change occurs in a series of small steps and to recalc
the two excess elasticities after each step

Introducing intermediate goods

The analysis up to this point has been based on the implicit assump
that all commodities are final consumption goods produced entirely f
original factors of production This assumption was very commo
theoretical work in the field of international economics until the mid-19
when a framework for handling international trade in intermediate g
was developed.

The new theory—the theory of tariff structures—utilizes an ex
demand and supply model in which the import supply curves are assu
to be infinitely elastic This modification of the model to allow for tr
in inputs has been analysed in detail elsewhere and will not be discus
in this paper [2] The generalized version of the excess demand and sup
model presented above does, however, prompt a comment on one of
assumptions underlying the tariff structure model.

The comment concerns the assumption that all import supply curves
infinitely elastic In utilizing this assumption, the theory of tariff structu
followed an already common practice—its popularity stemming in la
part from the fact (noted above) that it simplifies the mechanics of e
mating the change in the level of imports caused by a change in the tar
However, whereas the assumption is largely a convenience as far as t
conventional approach is concerned, it is virtually indispensable to t
tariff structure model [3]

By giving 'equal time' to the expression for the ϵ_x facing the tari
changing countries, the paper draws attention to the fact that in ma
instances the assumption of infinitely elastic import supply curves may
unduly restrictive This risk arises, in particular, when the tariff chang
involve major trading countries and trading blocs. Some insight into t
restrictiveness of the assumption in a given situation would be gained

[1] This assumption is less reasonable as regards the elasticity in the third term of ea
equation (ϵ_c and η_c)—elasticities which are themselves excess elasticities While in princi
they could be handled in a way similar to the stepwise recalculation of η_m and ϵ_x suggest
below, such a refinement is likely to be impractical in empirical work

[2] See [4], [5], [11], and [12]

[3] Corden has observed, for example, that ' . when the elasticities for inputs are le
than infinite, the effective-protective-rate concept strictly interpreted appears to break dow
[5, p 236] See also [16], [17], and [23]

researchers who wished to take advantage of it would provide estimates of ϵ_x, or, at least, of the values of the ratios in equation (4)

Use of the model in tariff negotiations

The great majority of people continue to view tariff reductions as concessions and for this reason tariff negotiations invariably involve the issue of 'balance of advantages' or 'reciprocity'. This was no less true of the Kennedy Round than earlier tariff negotiations [1]

Ernest Preeg, a participant in the Kennedy Round negotiations from mid-1963 until their conclusion in June 1967, notes that four principal measures of the admittedly complex concept of reciprocity emerged during the negotiations: average depth of cut, trade volume offered for concession, loss of duties collectable, and projected trade impact [2] Regarding the fourth measure he observes:

> The impact of tariff and other concessions on future trade might be considered the main criterion of reciprocity. It is not possible, however, to calculate the effect precisely. Assumptions must be made as to supply and demand elasticities, which can vary from commodity to commodity and country to country. Differing interpretations of the value of tariff protection complicate the calculations further. These projections, when carried out during the Kennedy Round, were primarily for internal use, although statements about qualitative differences in the value of offers rested on some kind of assumption as to the expected trade-creating effect of the concessions [21, p 133.]

It is very likely that elementary versions of the excess demand and supply model have been used in the past to provide tariff negotiators with estimates of the 'impact of tariff and other concessions on future trade'. Four of the references above (p. 80, nn 2, 3) are dated 1957 or earlier and certainly were known to the people who had the responsibility of providing estimates for tariff negotiations during the postwar period It is also likely that they were aware of Barend de Vries's 1951 study of U S import demand elasticities, based on estimates provided by the U.S. Tariff Commission of the trade effects of hypothetical tariff changes, and the support lent to the excess demand and supply approach by his results [3]

[1] The proposal that the developed countries grant tariff preferences to the exports of the less-developed countries involved an essentially identical issue—in this instance the concern was with the issue of an equitable 'sharing of the burden' among the developed countries

[2] [21, pp 132–3] Preeg's discussion of the reciprocity issue includes a quotation from the final report on the United States negotiations The quotation reads in part as follows "' . in the course of the negotiations, numerous other factors [in addition to the value of trade covered by the concessions and the depth of tariff reductions] were considered in evaluating the balance of concessions—the height of duties, the characteristics of individual products, demand and supply elasticities, and the size and nature of markets . "' [21, pp 130-1].

[3] [6.] After calculating the implicit import demand elasticities for 176 commodities, de Vries calculated a weighted average elasticity for tariff reductions of $-2\ 23$ for the aggregate When he compared the average elasticity for the group of commodities whose import-consumption ratios were below the average ratio for all 176 commodities (27 per cent)

Useful as this approach may have been in the past, it was limited by tl elementary nature of the models and by the use of *a priori* assumptioı about the value of η_m or ϵ_x The generalized version of the excess demaı and supply model developed in this paper offers future trade negotiato the opportunity to work with estimates of the trade effects of trade coı cessions whose accuracy—while still far from perfect—generally will repr sent a considerable improvement over that of previous estimates.

University of Waterloo, Ontario

REFERENCES

1. BALASSA, B., *Trade Liberalization Among Industrial Countries*, New York, 196'
2. —— and KREININ, M E., 'Trade liberalization under the "Kennedy Round' the static effects', *Rev Econ. Stat.*, May 1967.
3. —— *et al*, *Studies in Trade Liberalization*, Baltimore, 1967.
4. BLACKHURST, R., 'A model for estimating the impact of tariff manipulation o the volume of imports', *Staff Research Studies*, U.S Tariff Commission, Washing ton, D.C. 1972.
5. CORDEN, W. M., 'The structure of a tariff system and the effective protectiv rate', *Jour. Pol. Econ*, June 1966.
6. DE VRIES, B A., 'Price elasticities of demand for individual commoditie imported into the United States', *IMF Staff Papers*, Apr. 1951.
7. FERGUSON, C. E, and POLASEK, M., 'The elasticity of import demand for raw apparel wool in the United States', *Econometrica*, Oct. 1962
8. FLOYD, J. E., 'The overvaluation of the dollar a note on the internationa price mechanism', *Am. Econ. Rev*, Mar. 1965.
9. HARBERGER, A., 'A structural approach to the problem of import demand', *Am. Econ. Rev.*, May 1953.
10. JOHNSON, H. G., *Money, Trade and Economic Growth*, Cambridge, Massachusetts, 1962.
11. —— 'The theory of effective protection and preferences', *Economica*, May 1969.
12. —— 'The theory of tariff structure with special reference to world trade and development', in *Trade and Development*, Geneva, 1965.
13. KOJIMA, K., 'Trade preferences for developing countries · a Japanese assessment', *Hitotsubashi Jour. of Econ.*, Feb. 1969.
14. KREININ, M. E., 'Effect of tariff changes on the prices and volume of imports', *Am. Econ. Rev*, June 1961.
15. LEAMER, E. E., and STERN, R. M., *Quantitative International Economics*, Boston, 1970.
16. LEITH, J. C., 'Substitution and supply elasticities in calculating the effective protective rate', *Quart. Jour. Econ.*, Nov. 1968.
17. —— 'Substitution and supply elasticities in calculating the effective protective rate: reply', *Quart. Jour. Econ.*, Feb. 1970.
18. MACDOUGALL, G. D A., *The World Dollar Problem*, London, 1957.
19. ORCUTT, G. H, 'Measurement of price elasticities in international trade', *Rev. Econ. Stat.*, May 1950.
20. PIQUET, H. S., *Aid, Trade and the Tariff*, New York, 1953.

with the average elasticity for the group whose ratios exceeded the average, he found that the former exceeded the latter— $-3\ 13$ versus $-1{\cdot}77$, the difference being statistically significant The estimates in the Tariff Commission study were made by commodity experts It is improbable that any of their estimates were based on an explicit excess demand and supply model.

21. Preeg, E. H., *Traders and Diplomats*, Washington, D.C., 1970.
22. Stern, R. M., 'The U.S. tariff and the efficiency of the U.S. economy', *Am. Econ. Rev.*, May 1964.
23. Wood, G. Donald, Jr., 'Substitution and supply elasticities in calculating the effective protective rate: comment', *Quart. Jour. Econ.*, Feb. 1970.
24. Yntema, T. O., *A Mathematical Reformulation of the General Theory of International Trade*, Chicago, 1932.

EXPORTS AND ECONOMIC GROWTH IN WEST MALAYSIA

By J. T. THOBURN[1]

IN the late nineteenth and early twentieth centuries many countries experienced rapid rises in primary product exports.[2] In most cases, these exports remain important to the present day in terms both of their absolute volume and value and of their contribution to national product. Some of the countries, such as Canada and Australia, now have high levels of *per capita* income. Others experienced much less economic growth and are still regarded as 'underdeveloped', although within the 'underdeveloped' group many differences exist in relative growth performance. It has been contended that rapid export growth in poor countries has led to the existence of export industry 'enclaves', which have not, and cannot, spread development into the rest of the domestic economy.[3] Such a view has implications both for the interpretation of past experience and for the formulation of present policy.

The relation between exports and economic growth in poor countries can be examined in aggregative terms using cross-sectional studies of a large number of countries. Another approach is the case study of individual exports in a particular country.[4] What case studies lose in generality they gain in depth. This paper summarizes the results of a case study of West Malaysia, concentrating on the effects of tin and rubber exports.

[1] This paper summarizes the results of a research project financed mainly by the Overseas Development Administration of the Foreign and Commonwealth Office, London, and submitted to the University of Alberta in Canada as a doctoral thesis in 1971. Of course, ODA is not responsible for any view expressed here.

I am grateful for very helpful comments to Professor P. T. Bauer of the London School of Economics, to Mrs. F. Stewart and Mr. P. P. Streeten of Queen Elizabeth House, Oxford, and Professor H. F. Lydall of East Anglia. I should like also to thank my thesis supervisor Professor T. L. Powrie for his help and encouragement throughout the project and the Faculty of Economics and Administration in the University of Malaya, Kuala Lumpur, for its hospitality during my two visits to Malaysia in 1969 and 1970. An earlier version of this paper was circulated as a University of East Anglia Economics Discussion Paper (No. 1, Mar. 1972), and a later draft was presented at the East Anglia faculty seminar.

[2] These countries are sometimes known as the 'periphery'. The contrast is with the countries of the 'centre', such as the United Kingdom, France, and Germany, which had already achieved a substantial degree of industrialization and growth.

[3] The most influential statement of this view is Singer (1950). Singer suggests that for a number of economic, social, and institutional reasons the favourable multiplier-accelerator effects of export growth bypass the host economy. He opposes the development of primary product exports because he feels that industrial development offers more growth effects and because of the supposed secular deterioration in the terms of trade of primary products. Later work has concentrated on assessing the mechanisms through which trade can stimulate growth. For surveys of this literature see Watkins (1963), Meier (1968, pp. 214–54), and Thoburn (1972c).

[4] For a cross-sectional study see Emery (1967). For case studies see Lim (1968) on Ceylon, Baldwin (1966) on Zambia, and Levin (1960) on Peru and Burma.

It is hoped also that the study will shed light on the current controversy about the value of the concept of 'linkages' in project appraisal [1]

I. Theories of trade and growth, and the choice of West Malaysia as a case study

Export sector effects on development are seen in this study as working through the disposition of export income flows and through externalities Essentially this is a theory of capital formation involving a disaggregated multiplier-accelerator mechanism [2] Growth of an export not only offers investment opportunities in the industry itself, but also in industries supplying inputs to the export sector, in industries using the export as an input, and in industries producing consumer goods for factors of production employed in exports Such additional investment opportunities, or 'linkages' in Hirschman's terminology [Hirschman, 1958, pp 98–119] are accelerator effects, although they are often large and discontinuous so that an accelerator of the usual macro-model type cannot be quantified Of course, the fact that an investment opportunity exists does not mean that it will be taken up by local suppliers In the analysis which follows, the term 'linkage' is applied only where local supplies have actually appeared. Linkages are seen as a long-run phenomenon, involving new investment rather than short-run changes in output

The relevance of linkage effects depends much on how the export in question was developed. Where the industry was set up by means of foreign investment, as has been often the case in practice, the linkage analysis is quite appropriate The greater proportion of intermediate to final purchases and of intermediate to final sales the more chance have local people to participate as capitalists These opportunities are in addition to any which may exist for direct participation in the industry. These linkages are especially important since foreign capital imports have often been associated with the large-scale importation of foreign labour,[3] thus reducing the possibilities of the foreign capital raising *per capita* output Where exports develop through domestic investment, the value of linkages becomes less clear An industry with a high degree of vertical disintegration (i e. with a high degree of backward and forward linkages) does not offer any greater opportunities for domestic capital formation than one without these attributes, except in the trivial sense that the degree of disintegration may be associated (fortuitously) with a higher

[1] See in particular Stewart and Streeton (1972) The controversy is discussed in Section I below

[2] This description is used by Watkins (1963), p 145

[3] Thus the rubber estate industry in Malaya in the early years of this century developed through the importation of Indian labour, and tin mining was associated with a large-scale inflow of Chinese from 1870 to the late 1920s. Chinese are now about a third and Indians 10 per cent of the population of Malaysia.

capital–output ratio Of course, vertical disintegration does allow an industry of given value of sales to be established with a smaller initial investment, so that further investment can take the form of substituting domestic production for previously imported inputs. Also, linkages may be desirable if the linked industries use a technology different from that of the export good This aspect is discussed below.

The linkage analysis relies on an assumption that savings depend on investment opportunities If they do not, then linkages involve merely a reallocation of existing investment resources This question is an empirical one, difficult to determine in practice, but in the early stages of a country's development it seems hard to underrate the value of investment opportunities being offered to domestic entrepreneurs However, it is interesting to note in this context that the influential Little–Mirrlees method of project appraisal minimizes the importance of linkages In Little and Mirrlee's view it is the supply of savings, not investment opportunities, which is a constraint on development. Moreover, the extent of linkages may not differ significantly between projects, and industries developing through forward linkage could in most cases be supplied equally well by imports [1] It will be shown in this study that tin and rubber do differ substantially in their generation of linkages, and that there is at least circumstantial evidence that many of the linkages were not at the expense of other investment

Because of doubts about the validity of 'linkages', some writers have put more stress on exports introducing new technology into a poor country [2] Yet 'new' or 'advanced' technology is an elusive concept New techniques could raise the rate of return on capital in an established industry, but if the export industry had not previously produced for local use then the technology would be 'new' by definition. What appears to be meant in much of the literature is that new technology may involve mechanization, either in the export industry itself or in linked industries, familiarity with which facilitates further industrialization.[3]

As well as offering pecuniary externalities in the form of linkages, an export industry may generate unpriced externalities.[4] One of the most

[1] See Little and Mirrlees (1968), especially pp 209–19, and Little and Mirrlees (1972), especially pp. 165–6 For an attack on Little and Mirrlees's treatment of linkages see Stewart and Streeton (1972) Little and Mirrlees's work is concerned with industrial projects, but it is widely applied to primary product projects also. Some empirical evidence about the dependence of saving on investment opportunities is discussed in Wolf and Sufrin (1955), especially pp 11–13 It should be noted too that the Little–Mirrlees method is concerned mainly with the analysis of projects marginal to the economy, whereas the introduction of, say, the rubber industry might alter many of the conditions in the rest of the economy which Little–Mirrlees would assume constant.

[2] See especially Baldwin (1963)

[3] See, for example, Solo (1966), pp 486–7.

[4] For a discussion of externalities see Scitovsky (1954). Scitovsky distinguishes between pecuniary and technological externalities, the latter being the unpriced costs and benefits

important of these is the development of a skilled labour force, which can be used by industries which grow up subsequently in the country

West Malaysia is an appropriate choice as a case study for a number of reasons[1] After Japan, it is one of the richest countries in Asia[2] and exports constitute nearly half of the national product[3] Thus the possibility exists of a connection between exports and growth The two main export industries, tin and rubber, include almost the whole range of export sector types discussed in the literature. Not only do they represent the two major types—minerals and agricultural exports—but each has both substantial foreign and domestic sectors Moreover, rubber is split into a sector based on an estate system and a sector based on peasant smallholdings Of course, the findings of a single case study are not enough to confirm or refute existing views on trade and growth Nevertheless, if it can be shown that in Malaysia exports have had a substantial effect on development, then the simplistic but influential view of export enclaves is brought into question and in any case light will be shed on the workings of the trade-growth mechanism in practice

II. Methods of analysis of the economic effects of export growth

The first stage in the analysis, carried out in Section III, is to split export sector payments into their various categories, and to determine the local and foreign component for each category. Multiplier effects work through final demand payments by factors of production employed in exports (and by government expenditure out of taxes on exports). The proportion of value-added paid to locally resident factors (i e workers and domestic recipients of profits), and the Malaysian government, indicates the average propensity for export income to accrue to potential local spenders In Section V an attempt is made to decompose this relationship into an autonomous element (which nevertheless can change over time), and a purely multiplier element, whereby changes in export income generate domestic factor income changes. The proportion of such local factor income spent on local products, also estimated in Section V, constitutes a second round of multiplier effects, and investment in these products (i e. final demand linkages) can thereby be generated through an accelerator The local content of government expenditure is not examined

of an investment. Both the causes and effects of such unpriced externalities as labour training and the generation of entrepreneurial ability have important economic aspects Hence the term *unpriced* externality is used here in preference to *technological* externality

[1] East Malaysia (Sabah and Sarawak on the island of Borneo) is excluded because few data are available for that region.

[2] In 1966 gross national product at market prices in West Malaysia was U S $316 *per capita*, compared with U S $190 for the whole of the East and South East Asia region and U.S. $120 for the region excluding Japan See United Nations (1970), p. 564

[3] In 1966 rubber exports were 45 per cent and tin exports 25 per cent of total West Malaysian export earnings. Other important exports were iron ore (4 per cent), palm oil and kernels (4 per cent), and timber (3 per cent) See Statistics Department (1970)

The proportion of export income calculated in Section III as being *retained initially in Malaysia* includes both value-added payments and payments for intermediate products Investment in intermediate products is 'backward linkage', and the types of industry which have developed through such linkage are discussed in Section IV[1]—which also looks at forward linkages and direct local export investment Section VI examines one of the most important unpriced externalities of export growth, the training of a skilled labour force.

For purposes of analysis each export industry is split into two sectors: tin dredging (the foreign sector) and gravel-pump tin mining (the local sector), and rubber estates and rubber smallholdings [2] The proportion of export income retained locally is particularly relevant for dredging and rubber estates, the foreign sectors, since there is a presumption that earnings may be remitted abroad. The local retention of earnings is also calculated for the local, gravel pump, mining sector to provide comparison with dredging A similar calculation is not possible for rubber smallholdings but only a very small proportion of their earnings is likely to be remitted abroad. For 1968, the latest year for which most statistics are available at the time of writing, a quite detailed breakdown of the payments structure is possible for the two tin sectors and for rubber estates [3] Since these export industries have been important since long before the Second World War, it is interesting also to determine whether their payments structures have altered significantly over time. For this purpose it is necessary to construct time series of the major inputs used in each sector and of

[1] Professor Bauer has suggested to me in conversation that the whole use of the 'linkages' concept is unnecessary to the analysis He stresses the importance of direct income-earning opportunities from exports and the responsiveness of local people to economic opportunity. Thus, presumably, investment by local people would have occurred whatever the pattern of linkages While this view contains an important element of truth—that responsiveness to investment opportunities is crucial—I feel, nevertheless, that an export industry which generates obvious local investment opportunities in other sectors is more conducive to further growth of the economy than one which does not. This is also important historically in Malaysia for the reason that local investment in the export industries themselves was made difficult by the various tin and rubber restriction schemes of the interwar period It must also be stressed that linkages are as much an economic as a technological concept in that the choice of production technique is likely to depend, to some extent at least, on factor prices and availability, and that there can be 'bad' as well as 'good' linkages See concluding paragraph of Section IV

[2] In 1968 61 per cent of West Malaysian tin output came from gravel-pump mines and 20 per cent from dredges, while gravel-pump mines accounted for 58 per cent of mining employment and dredges 31 per cent See Mines Department (1968) In 1968 rubber estates had 30 per cent of total West Malaysian rubber acreage and accounted for 54 per cent of rubber output All other acreage and output was in the hands of smallholders. See Statistics Department (1968c)

[3] There is not space here to discuss for each component the methodology used to split local from foreign payments Also, it is not possible to present in any detail the time series of inputs and payments used to make comparisons with earlier years, nor their sources and methods of construction Any reader interested in these details is invited to contact the author at the University of East Anglia.

payments made to the government, and this also is carried out in Section III

This study does not consider the terms of trade of Malaysia, nor the effects of export income instability, except in so far as they affect the export industries' payments structures

III. Payments structure of tin and rubber

Table I shows the structure of payments made out of the current export income of the tin-dredging sector and illustrates the approach to be used here. Information on wages and materials is available from the West Malaysian *Census of Mining Industries* Wages can be broken down into foreign versus local payments by taking foreign payments as those to expatriate mine managers and engineers,[1] the numbers of which can be found in government publications, and multiplying the numbers by an annual salary assumed on the basis of personal inquiries in Malaysia The breakdown of materials purchases given in the *Census* allows local payments to be identified, the principal one being electricity Information on other operating costs, depreciation, and U K expenses, was obtained from a sample survey of annual reports of tin-dredging companies in Malaysia and the U K [Thoburn, 1971, pp 184–8] Export duty payments are from government reports Tax and total dividend payments and profits are available from a manual of tin companies' balance sheets produced annually by a London firm of stockbrokers [Messrs Zorn and Leigh-Hunt, 1969*b*] although reworking these data into a usable form requires the summation of profits, etc., company by company The information of profits provides a useful check on the cost information, since pre-tax profits should in principle equal the residual left from export income after costs and export duty payments have been accounted for. Dividend payments were split into their local and foreign components on the basis of the relative shares held by overseas and Malaysian residents Information on shareholdings in individual companies was obtained by means of a postal survey of U K -incorporated dredging companies, while data on Malaysian-incorporated firms were taken from an existing study [Yip, 1968] The holdings were then weighted by the current market value of the issued capital of the company concerned in order to arrive at an over-all figure for the Malaysian ownership of capital[2]

The striking feature of Table I is the very high proportion of locally retained income, equal to nearly three-quarters of the annual value of tin sales Much of this local retention is due to government policy—over a third of total export income is paid to the Malaysian government in the

[1] This is to allow for the fact that they may have such high propensities to import that their incomes would give a misleading picture of their likely local expenditure

[2] For further details see Thoburn (1971), pp 468–9, and Thoburn (1972*b*)

form of tax and export duty. Also noteworthy is the fact that approximately a third of dividends are paid to local shareholders, and that most materials are bought locally. Thus the only important outflow is foreign dividends, which account for some 15 per cent of export income.

It is interesting to work back into time to trace whether the present payments structure is significantly different from the past. Data are not available to produce a complete breakdown of the sort shown in Table I

TABLE I

Local and foreign payments in tin dredging, West Malaysia, 1967

(Items as percentages of annual value of output)

Payments	*Total*	*Local*	*Foreign*	*Unallocated*
Costs				
Wages	14·8	13·3	1·5	..
Materials	13·7	8·2		5·5
Other operating costs	6·3			6·3
Depreciation	6·5	6·5		.
U.K. expenses	1·7		1·7	
Other costs	0·5			.
Total costs	43·5	28·0	3·2	11·8
Allocation of gross profits				
(gross profits = 100 − total costs + 'other proceeds' of 4·3% = 60·8%)				
Export duty	14·7	14·7		..
Malaysian tax	20·7	20·7		.
U.K. tax	1·1		1·1	..
Total allocation of export income (value of sales + 'other proceeds')	104·4	72·5	19·6	12·3

SOURCES AND NOTES (1) For a detailed description of sources and methods see Thoburn (1971), pp 177–98, and especially Table IV–20 on p. 192. (2) 'Other proceeds' consist of income received by the companies from sources other than the sale of tin ore. Usually these sources are income from holdings of securities. (3) The treatment of depreciation as a local 'payment' may seem contentious as only a quarter of dredge capital expenditure is made locally (see Section IV). However, there is no *a priori* reason to suppose that firms in Malaysia remit abroad earnings retained as depreciation allowances. The local content of capital expenditure is best treated as a separate issue.

for any year earlier than 1964. Nevertheless, it was possible to construct relatively reliable time series of wage payments, export duty payments, and purchases of electricity (the main material input).[1] No major changes have occurred in these inputs as percentages of the value of output,[2] hence the proportion of pre-tax profits in gross output will have been relatively

[1] For details see Thoburn (1971), pp 198–213 and 470–81. This reference also covers gravel-pump tin and the period 1910–39.

[2] Except for falls in the percentages in times of very high tin prices, especially during the Korean boom. See Section V for a discussion of the relationship between fluctuations in export earnings and payments to factors employed in the export industries.

constant. The disposition of these gross profits will have changed, however First, the Malaysian tax on company profits has been at its present 40 per cent level only since 1959, and was first imposed (at 20 per cent) in 1948. Second, available evidence suggests that the present substantial Malaysian holdings of dredging shares did not exist before the mid-1950s [Yip, 1968, pp. 71–8]. Thus before 1948 it is likely that an additional 30 per cent points of tin revenue was available for remission abroad

For the period before the Second World War a surprising range of statistics is available on dredging inputs, although the quality of much of this information is suspect It was possible to compile a wage series for the 1910–39 period (though of less reliability than the postwar one),[1] a series for electricity purchases for some years in the 1930s and a complete export duty series for the whole period, together with details of smelting and transport costs, and purchases of coal and wood fuel [2] The 1930s show differences from the postwar years Labour accounts for at least 5 per cent points less in total output value, while electricity purchases are only some 3 per cent of output value (though other fuels were more important then than now) In the absence of any local company tax and any significant local shareholdings, and with relatively constant export duty, it is likely that the potential outflow of export income was at least 50 per cent, and possibly more

Table II performs a similar exercise for gravel-pump tin. Since this sector is almost entirely locally owned there is no presumption that earnings will be remitted abroad The main interest in a breakdown of the input structure therefore is, first, to provide a comparison with dredging, and to indicate the existence of linkages On the assumption that profits (after duty and local tax) are retained locally, apart from a small dividend outflow from the minority of foreign firms in the sector, over 85 per cent of export income is retained Materials purchases are mainly of electricity and diesel, both locally produced, though diesel has a high indirect import content

Working back the main input series into the past is possible only to a limited extent in gravel pumping, since no statistics are available of diesel consumption before 1964 However, the wage bill is relatively constant

[1] The postwar series was based on employment figures pro-rated by an average wage calculated from Malaysian Ministry of Labour sources, which gave a detailed breakdown of wages paid to different categories of worker together with details of the relative importance of each category in the total work force The prewar statistics were based on 'average' wages quoted in government sources.

[2] In fact, series for coal and wood fuel consumption are available only for the tin-mining industry as a whole Since residual profits are likely to be retained in the country by gravel-pump mines, and remitted by dredging companies, it is necessary to know the breakdown of materials purchases by sector before the outflow of payments can be assessed Also, no output figures by sector are available for the period before 1928, so that the payments outflow can be calculated effectively only after that date See Thoburn (1971) pp 205–13.

(at around 20–5 per cent of output), as is electricity and export duty though with some larger fluctuations in years of especially high or low tir prices [Thoburn, 1971, p 202]. Thus, other than for the fact that diese was imported before 1963 [Thoburn, 1971, p. 226] (which would have increased the payments outflow overseas by about 10 per cent points), the payments structure of gravel-pump tin has remained relatively stable since the Second World War In the 1930s, too, electricity purchases appear to have accounted for about the same proportion of output as after the war

TABLE II

Local and foreign payments in gravel-pump tin mining, West Malaysia, 1967

(Items as percentages of annual value of output)

Payments	*Total*	*Local*	*Foreign*	*Unallocated*
Costs				
Wages (including free food)	23 7	22 7	1 0	
Materials	23 1	12 5	3 4	7 2
Total costs	46 8	35 2	4 4	7 2
Allocation of gross profits				
(100 – total costs – 53 2%)				
Export duty	14 7	14·7		
Foreign dividends	2 0		2 0	
Other payments	36 5	36 5		
Total allocation of export income	100 0	86 4	6 4	7 2

SOURCES AND NOTES (1) For full details see Thoburn (1971), pp 192–8. (2) Approximately 10 per cent of the output of gravel-pump tin is produced by foreign companies Assuming their share of total profits after export duty to be 10 per cent also (i e 10 per cent of 53 2 – 14 7 = 3 81), and assuming they pay 40 per cent company tax on their profits, the maximum foreign dividend payable would be approximately 2 per cent, as shown (3) Other payments includes both local tax payments and local profits

For tin mining as a whole in the 1930s, diesel consumption (an outflow) was 2 per cent to 4 per cent of output value, and diesel would have been used by both sectors

Table III gives a breakdown of local and foreign payments made by rubber estates out of rubber export income Again, space does not permit a full description of the data sources or the methods used to achieve the breakdown The same principle was used as in dredging of calculating gross profits first as a residual after labour and materials inputs had been quantified, and checking the result against figures for company profits obtained directly from company returns Company returns also provide information on tax and dividend payments, etc. Table IV shows the data on profits, taxes, and dividends compiled from the reports of U K and

TABLE III

Local and foreign payments of the rubber estate sector, West Malaysia, 1967

(Items as percentages of annual value of output)

Payments	*Total*	*Local*	*Foreign*	*Unallocated*
Costs				
Wages and labour benefits	46 0	39 4	6 5	
Materials	12 4	6 2	6 2	
Total costs	58 4	45 6	12 7	
Allocation of gross profits (100 − 58 4 + 'Other receipts' of 4 7 = 46 3%)				
Export duty	3 5	3 6		
Research cess	1 6	1 6		
Replanting cess	7 4	7 4		
Malaysian tax	9 9	9 9		
Dividends	13 4	2 7	10 7	
Net replanting and new planting expenditure	4 4	4 4		
Other payments	6 1			6 1
Total allocation of export income	104 7	75 1	23 4	6 1

SOURCES AND NOTES (1) For full details see Thoburn (1971), pp 292–305, especially p 304 (2) Other receipts are mainly income from investment holdings (3) Much of the breakdown of materials items is from Rubber Research Institute of Malaya (1969)

TABLE IV

Profits, taxes, dividends, and planting expenditure of public rubber companies in Malaysia, 1967–8

	(i) *Number of companies*	(ii) *Annual value of output (M$ mil)*	(iii) *Annual value of rubber output (M$ mil)*	(iv) *Profits before tax and after cesses and duty as % of (ii)*	(v) *Tax as % of (ii)*	(vi) *Dividends as % of (ii)*	(vii) *Net new and replanting expenditure as % of (ii)*
1 *U K -incorporated companies*							
Rubber companies	48	101 4	101 4	22 7	6 3	10 9	7 8
Multicrop companies	10	269 7	185 9	32 4	12 4	15 0	4 0
Rubber and multicrop companies	58	371 0	287 4	29 7	10 7	13 8	5 1
2 *Malaysia-incorporated companies*							
Rubber companies	18	31 6	31 6	26 1	6 9	10 5	2 2
Multicrop companies	4	50 7	32 3	25 6	5 6	11 9	0 5
Rubber and multicrop companies	22	83 0	63 9	25 8	6 1	11 4	1 1
3 *Malaysia and U K -incorporated companies*	80	454 2	351 3	29 0	9 9	13 4	4 4

SOURCES AND NOTES (1) Calculated from Zorn and Leigh-Hunt (1969*a*) (2) Tax is almost entirely Malaysian tax (3) Net new planting and replanting expenditure refers to expenditure made in addition to that financed by refunds of the replanting cess

Malaysian-incorporated rubber companies.[1] Profits before tax are shown to be equal to nearly 30 per cent of the 1967–8 output value, to which figure must be added approximately 10 per cent points to take account of export duty and cesses. The resulting 40 per cent accords well with the profit figure calculated in Table III as a residual

Wage payments in rubber are of much greater importance than in tin, and the share of profits in rubber output is less. Payments of Malaysian taxes, cesses, and export duty retain about 20 per cent of export income within the country, but a low proportion of local shareholders means that most dividends are repatriated Over all, however, the proportion of retained earnings is high Few changes have occurred in wages as a proportion of output (except in years of very high prices) since the Second World War, though before 1948 no payments of Malaysian tax would have been made, thus increasing the potential outflow Prewar figures indicate that the wage bill was also about 40 per cent of output value in the 1930s although large fluctuations in the proportion occurred with the large price changes during that period [2]

The picture which emerges from these results is one in which the export industries retain a high proportion of income in Malaysia for the first round of the multiplier This destroys the idea that they are 'enclaves' in the obvious sense of immediately remitting most of their income abroad. However, the initial retention of export income, although important, is only the first stage in the process by which exports may stimulate growth. An examination is also necessary, first, of the linkages generated by each sector through intermediate demand, and of the extent to which local entrepreneurs have participated directly in export production, second, of final demand linkages and third of the extent to which the quality and earning power of workers in the sector has been improved, and unpriced externalities generated

IV. Export sector forward and backward linkage effects and direct participation

Tables I and II show that materials purchases account for some 14 per cent and 23 per cent of total payments made from export income in the dredging and gravel-pumping tin sectors, respectively. Over 80 per cent of the materials purchased by the industry are fuel [Statistics Department 1968*b*]. Electricity is consumed in large quantities by both sectors, and diesel oil mainly by gravel pumping.

There seems little doubt that the large-scale generation of electricity in Malaysia, first established in 1928, owes at least its early existence to the

[1] Almost all public rubber companies in Malaysia are incorporated either in the U.K Malaysia and U K -incorporated public rubber companies account for 80 per cent of West Malaysian foreign rubber estate output. See Thoburn (1971), p. 303.

[2] See Thoburn (1971), p. 312 for these time series.

mining. In the 1930s over 80 per cent of the units generated by public utilities were sold to mines. This figure was still 70 per cent in 1950, though the emergence of other large consumers had reduced it to 40 per cent in 1967.[1] It is certainly a plausible proposition that the presence of a large-scale electricity supply industry, established to meet the needs of tin mining, should have facilitated the development of manufacturing in Malaysia by providing a source of cheap power.

The local production of diesel oil and other petroleum products dates from 1963–4, when two refineries were established by Esso and Shell in Port Dickson in Negri Sembilan Tin mining is the largest single industrial consumer of petroleum products, but it is difficult to see the relation as one of direct economic linkage since mining buys only 10 per cent of the petroleum output Nevertheless, it is possible that the existence of such a major consumer may have influenced the investment decision of at least one refinery

A third backward linkage of tin was coal mining. Coal was mined by Malayan Collieries Ltd at Batu Arang in Selangor from 1915 to 1960 No figures are available for postwar tin consumption of coal, but during the 1915–39 period tin mines bought between a third and a half of the coal output, the railways being the other main consumer Tin consumption of coal declined as steam power was replaced by electricity and diesel At their peak output in 1929, Malayan collieries employed 4,000 workers [Mines Department, 1955]. However, it was financed by European tin-mining interests, and thus represented backward integration of the foreign sector into intermediate production rather than genuine local participation

The beginning of railway development in the country was closely connected with tin mining The first railways, built in the 1880s, were short east–west lines connecting mining centres with sea ports By 1910 a line had been built between the various mining centres connecting Prai in the north with Jôhôre Bahru in the south, and was extended over the causeway to Singapore in 1923 This still constitutes the basic west-coast railway system of the present day [2]

Finance for early railway construction came mainly from the current revenues of the Federated Malay States,[3] while over a third of this revenue came from the tin export duty [4] However, there was little connection

[1] Electricity Department (1937), Central Electricity Board (1950 and 1967) As Table V shows, these figures overestimate the *value* of electricity output sold to mines since mines receive large discounts on the price per unit as large consumers

[2] For an account of railway development see Railways Department (1935)

[3] Lim Chong Yah (1967), p 275, calculates that from 1884 to 1937 (the period during which almost all the present system was built) over 75 per cent of railway construction was financed from current F M S revenue

[4] From 1884 to 1910—the peak of railway construction—tin export duty constituted 36 per cent of total F M.S revenue For details of the sources of this figure (mainly government annual reports) see Thoburn (1971), p 232

between tin mining and railway building in the sense of direct econon linkage In 1916, the earliest year for which statistics are available, t transport of tin ore generated only 2 per cent of total railway revenue a 5 per cent of railway goods revenue in the Federated Malay States.[1] T early railway derived their revenue from the general development resulti from tin mining and later from rubber, rather than from the transport ore (or rubber) itself

Virtually all tin mined in Malaysia is smelted in that country by t smelting firms located in Penang Both firms have a history dating ba to the nineteenth century [2] The only other forward linkage is the lo manufacture of pewterware, which is sold particularly to tourists, a takes a negligible amount of tin output

Table III shows that 12 per cent of rubber estate export earnings spent on materials. This figure drops to approximately 8 per cent if rubl smallholdings are included [3] The rubber industry's main materials p chases are of chemicals (fertilizers, weedicides, coagulating acid, and ar coagulants) The agricultural chemical industry is of quite recent orig mainly comprising local branches of multinational companies such Imperial Chemical Industries and Dow, set up behind a tariff wall as p of an import substitution policy. In addition, minor items of rubber est equipment such as tapping knives and churns to carry liquid latex made locally, mostly by small-to-medium-size Chinese businesses

There are two forward linkages from rubber. the off-estate process sector and the manufacture of rubber products Off-estate processing essentially a satellite industry, in Hirschman's terminology [Hirschm 1958, pp 102–4] [4] The manufacture of rubber goods in Malaysia has b carried out since long before the Second World War. A wide range products is made, including rubber shoes, tyres, and foam rubber produ The industry, however, takes less than 2 per cent of the total W Malaysian rubber output [Statistics Department, 1968c].

All the linkages so far described have operated through current p chases and sales of intermediate products. Linkages can also oper through purchases (or sales) of capital goods. It can be shown that 60 cent of the value of a rubber-processing factory, 25 per cent of a dre

[1] Railways Department (1916) The figures for rubber were 5 per cent and 10 per respectively

[2] For further information on smelting, based mainly on interviews with the two panies, see Thoburn (1971), pp. 173–4 and 234–5.

[3] This is the 1965 figure from Department of Statistics (1965).

[4] Rubber processing must normally be carried out soon after the rubber has been tap It is unlikely that the shipping of rubber in improved form (which in effect mear naturally coagulated scrap) could ever have been considered seriously, in view of the price which such rubber would have fetched. In contrast, the shipping of tin or smelting overseas is quite feasible, and much foreign ore is smelted in the United King for example.

and 40 per cent of the equipment for a Chinese mine is usually purchased locally.[1] These capital payments have given rise to a substantial light engineering industry which produces iron and steel castings, structural steelwork, and a wide range of equipment. The development of the engineering industry, which dates mainly from the interwar period, is described in a separate case study [Thoburn, 1972*a*].

Table V gives an over-all view of export sector linkages and some information about their relative size and economic structure compared with the export industries themselves. The small employment in all rubber linkage industries compared with even the rubber estate sector alone is apparent. The non-wage value added (NWVA) figures per full-time worker (which can be taken as a measure of the flow of capital services in each industry) show the relatively low capital intensity of old established linkages such as rubber goods and engineering, compared with newer industries such as petroleum refining and chemicals. The high capital intensity of mining compared with domestic manufacturing and more especially with rubber is also apparent.

Direct local participation in the tin industry has been limited until recently to the gravel-pump sector which is controlled largely by Malaysian Chinese. Until 1962, when the (Chinese) Selangor Dredging Company was established, no dredging company had been set up on purely local (non-European) initiative. At present this company, together with a Chinese private company operating a small second-hand dredge in Selangor, are the only examples of completely local participation, although, as was mentioned in Section II, the shares of European dredging companies have been purchased in large numbers by local people.

In the rubber industry nearly half of the estate acreage is in the hands of firms at least 50 per cent owned by local people [Statistics Department, 1966] while all the smallholding acreage is locally owned.[2]

Finally, there is the crucial question of whether the linkages described above represent additional investment or merely reallocation of existing investment funds. Of necessity the evidence is circumstantial: one cannot know what investment would have been in the absence of linkage opportunities. Nevertheless, some pointers exist. One, if not *the*, major alternative investment opportunity for local people in the years before the Second World War, when many linkages first developed, lay in the export industries

[1] These figures are derived from a series of interviews in Malaysia from Oct. to Dec. 1970 with manufacturers of engineering equipment and with estates and mines. If field establishment costs are also taken into account, the proportion of local purchases in rubber rises to 70 per cent. See Thoburn (1971), pp. 237–47 for the tin capital breakdown and pp. 326–32 for rubber.

[2] In 1952, one of the last years for which ownership figures are available by race, 47 per cent of the smallholding acreage was owned by Malays, 41 per cent by Chinese, and 8 per cent by Indians. See Department of Statistics (1952).

TABLE V

Forward and backward linkages of tin and rubber, 1968, West Malaysia

	Percentage of value of linked industry's output bought by export sector (for backward linkages), or percentage of value of linked industry's output bought from export sector (for forward linkages)	*Employment full-time workers only)*	*Average annual wage (M$)*	*Average annu non-wage valu added per work (M$)*
Tin—backward linkages				
Electricity supply	21% (1965)	n/a	n/a	n/a
Petroleum refining	10% (1965)	404 (maximum)	12,000	100,000
Tin—forward linkages				
Tin smelting	95% (minimum)	1,200	n/a	6,000
Manufacture of pewter	n/a	n/a	n/a	n/a
Rubber—backward linkages				
Chemical fertilizers	35% (rubber estates only)	587	4,279	13,850
Rubber—forward linkages				
Off-estate processing				
—smokehouses	84%	979	1,244	3,501
—remilling and latex concentrating	82%	8,605	1,939	5,423
Rubber manufactures	15%	8,375	2,144	3,911
Comparisons of economic structure				
Gravel-pump tin mining		35,514	2,208	5,710
Tin dredging		10,720	3,048	11,310
Rubber estates		206,680	1,368	1,160
All manufacturing industry		120,807	2,209	5,024
All pioneer establishments		22,652	2,792	9,493
All non-pioneer manufacturing industry		98,155	2,075	3,993
Manufacture of industrial machinery and parts	35% (minimum)	3,947	2,049	1,858

SOURCES AND NOTES (1) Values of output of linked industries from Statistics Department (1965) electricity and petroleum, from Statistics Department (1968*a*) for pewter, fertilizers, off-estate proces rubber manufactures, and manufacturing Calculations based on export industry outputs from Stati Department (1968*b*) for gravel pumping and dredging, and from Statistics Department (1968*c*) for rul (2) Employment, wage, and non-wage value added figures from same sources with the following excepti petroleum refining based on data from Statistics Department (1968*a*), but see Thoburn (1971, p 228) discussion of the limitations of these data, smelting based on interviews in Malaysia with the smelting panies, Nov 1971, and calculations from Thoburn (1971, pp 234–5) (3) The proportion of the outp industrial machinery and parts sold to the export industries is calculated in Thoburn (1971, pp 416–17) is a backward linkage generated not only by tin and rubber, but also by oil palm Hence it is shown separ from the main body of the table

themselves Yet from 1922 to 1941 (and indeed to the present day in case of estates) the alienation of new land for rubber (and oil palm) largely banned [Bauer, 1948, p xiii] and tin production was restricted the early 1920s and the 1930s. [Thoburn, 1971, pp 482–3] Thus link investments provided an alternative to the other favourite local uses profits—private residential construction and the remission of profits China Also it should be remembered, if the experience of the Malays engineering industry is typical, that much investment in linked indust was financed from retained profits, and there may be limits to the mobi

of such funds to other sectors [Thoburn, 1972*a*].[1] However, it is possible that some of the linkages may have been 'bad' ones, in the sense that they represent high-cost import substitution fostered by government policy and unmatched by a sufficient excess of social over private benefit to justify, from the point of view of promoting growth, the departure from comparative advantage Chemical fertilizer production is the only obvious candidate Engineering, on the other extreme, seems an unambiguously 'good' linkage [Little and Mirrlees, 1972, p. 165]

V. Final demand linkages

The extent to which export sector factors of production spend their income within the country, and the type of market created by their expenditure, are important determinants of the size and type of investment opportunities offered and of the possibilities of diversified growth beyond the initial export base The aim of the analysis is to determine what are the consumption expenditures of export sector workers and smallholders,[2] to split these payments into their local and foreign components, and finally to examine the effects of these local expenditures.

The main source of information on consumption expenditure is the *Household Budget Survey of the Federation of Malaya, 1957–58*, published by the Malaysian Department of Statistics This gives detailed breakdowns of expenditure, commodity by commodity, for Malay, Chinese, and Indian households For each group information is given separately according to a number of income categories and according to whether the households are rural or urban Given the form of the data, the first step is to decide what proportion of the wage bill in each industry (or total export income in the case of smallholders) accrues to each racial group. Second, it must be decided into which income category each of the groups in each industry falls and whether they are rural or urban. The total income received by each category in each industry is then multiplied by the proportion of expenditure spent on each good in the household budget, in order to show the total demand for each good. This procedure is repeated for each category, and the resulting demands are summed to arrive at the total demand by the export sector for each good. Since the procedure used to decide whether these demands are met by domestic production or by imports does not work for expenditures by small individual groups (because, as will be shown, the group's consumption could be compared only with total West Malaysian imports of the product concerned), the consumption

[1] However, work by Drabble (1967), pp 58–9, indicates some mobility of capital at least between the export industries themselves (from tin to rubber)

[2] There is insufficient information on the incomes of export sector profit receivers to assess their consumption patterns.

demands are summed for the whole of the tin and rubber industries T
simplify the already lengthy calculation, goods on which no group spe
more than 1 per cent of its income were omitted This reduced the number
goods from two hundred to fifty, while reducing the coverage to 75 per ce
of consumption expenditure.[1] For most of the goods concerned, domest
production figures were not available, so that export sector consumptic
could only be compared with net imports (i.e. imports–exports). Worke
in the tin and rubber industries, and rubber smallholders, comprise rough
a third of the West Malaysian working population Assuming that the
ratio of dependants to workers is roughly similar to that of workers in oth
industries, and since they represent a cross-section of most racial groups ar
income categories in the country, it could be assumed, as an approximatio
that they would generate a third of the demand for most consumption item
Thus if export sector demand for a good is equal to a third (or less) of n
imports, it would imply that domestic demand was met entirely fro
imports. If export sector demand divided by net imports is 1, then it
likely that as much as two-thirds of domestic demand is met by domest
production Only 1·1 per cent of the value of consumption expenditu
covered was on goods for which the demand: net import ratio was le
than 1. Items for which no import statistics are available constitute
another 11 per cent, though 5 per cent points of that was for fresh fish, whi
are almost certainly locally caught Since a demand. net import figure of
means that imports are likely to meet a third of export sector demand ar
assuming that the 25 per cent of consumption expenditure which w
excluded from the calculation have a roughly similar local component, th
at least two-thirds of export sector consumption may be made locally

The composition of local consumption expenditure as well as its tot
size is important. Demands for goods beyond simple foodstuffs can provi
opportunities for growth of a diversified industrial base and may introdu
new technology into the economy In fact 65 per cent at least of total lo
consumption expenditure by the export sector is on food. Of the mar
factured food items, sugar (4 per cent of expenditure) is refined locally, a
various bakery products (1 per cent) and condensed milk (3 per cent) are a
locally made. Non-food manufactures include cigarettes (4 per cent), wh
expenditures on 'modern' services, such as cinema tickets and bus far
each are about 1 per cent of expenditure Thus the market created is
simple products. However, the results exclude managers' and executiv
consumption, for whom no adequate consumption data are available.[2]

[1] Details of how it was decided to split the wage bills, allocate groups to the var
income categories, etc, are given in Thoburn (1971), pp 357–67

[2] The *Household Budget Survey* covers incomes of up to only M$1000 a month, whi
well below that earned by most executives, who would earn an amount equal to ano
10–20 per cent of the wage bill

The possibility of final demand linkages depends on the proportion of export income paid to local factors of production (shown in Section III) and on the proportion of that income spent locally, calculated above. It is also interesting to determine the extent to which local expenditure reacts to *changes* in export income, to see the multiplier effects of such changes, which may give rise in turn to investment through an acceleration mechanism. An attempt has therefore been made to estimate a function relating domestic consumption expenditure to export income, on the assumption that total wage payments in each export industry can be used as a proxy for consumption expenditure.[1, 2] In the rubber estate sector over the period 1946-68 marginal propensities of 0·03 to 0·06 were calculated, using a simple linear equation, but these were statistically not significantly different from zero. Lagging the equation by one year did not improve the fit. There was, however, a large and statistically significant intercept in the function. In contrast, statistically significant results were obtained for both tin sectors yielding MPCs of 0·12 in gravel pumping and 0·08 in dredging.[3] These results indicate that marginal changes in export income from rubber (and to a lesser extent from tin) accrue largely to profit earners and the government, though the existence of a large intercept in the rubber function and a relatively constant wage share suggests that labour's share of export income catches up after some lag, which is not statistically identifiable. The government's share of changes in export income would equal the export duty and cesses (15 per cent on tin and approximately 10 per cent on rubber) and the 40 per cent share of tin and rubber profits, together with the small Development Tax and Tin Profits Tax. Thus the first-round expenditure in a multiplier process resulting from a change in export earnings would be small (and even smaller to the extent that export workers have a positive marginal propensity to import). The large intercept in the function shows that the substantial local expenditure made out of export income is of an 'autonomous' type, not fluctuating with export earnings.

[1] For details see Thoburn (1971), pp. 368–72.

[2] If export sector workers are assumed to save only negligible amounts, to spend very little marginal income on imports, and not to pay any marginal income tax (which is true in practice), and export sector local profit receivers are assumed not to consume out of their export income, then the marginal propensity to consume out of export income will depend primarily on the relationship between export earnings and total wage payments (and on expenditure out of taxes accruing to the Malaysian government).

[3] These figures should be seen against the fact that wages in the two tin sectors are approximately 24 per cent and 15 per cent of total output value, respectively. Thus the MPCs indicate that a given rise in tin output value was matched by a proportionate increase of about 50 per cent in the wage bill, compared with 10 per cent for rubber estates where wages are some 40 per cent of output value. The low MPCs accord well with the findings of an existing study (Harvie, 1964).

. Labour training

[Two aspects of labour training are important. The first is the proportion the labour force in each export industry which requires and develops ›our skills, and thereby improves its earning power. This is a direct and ar effect of export development on the *per capita* income of local people o participate in the sector as workers It can be seen from Table VI ıt ın gravel pumping and more especially dredging, a relatively high ›portion of the work force is skilled in comparison with rubber, and these

TABLE VI

Skilled labour coefficients in gravel-pump mines, tin dredges, and rubber estates, West Malaysia, 1968

	Average monthly wage (all workers) M$	*Propor-tion of work force skilled*	*Skilled workers per M$ million of output*	*Total workers per M$ million of output*	*Total workers per M$ million of investment*	*Skilled workers per M$ million of investment*
›vel-pump n mines	184	0 20	17 2	86 1	234 9	47 0
dredges	254	0 41	20 0	48 8	19 4	8 0
›ber estates	114	0 07	21 2	302 3	78 4	5 5
nanufacturing ıdustry	184	n/a	n/a	39 0	n/a	n/a

›URCES AND NOTES (1) See Thoburn (1971), pp 336, 337, and 340, and Statistics Department 8a) for manufacturing statistics (2) Skilled workers include artisans, supervisors, managers, but such workers as rubber tappers, whose skills are not transferable to other industries

'erences in labour equality are reflected ın the higher wages paid in tin ıing. The wage in gravel pumping is almost exactly the same as in laysian manufacturing [Statistics Department, 1968*a*] indicating use of nparable labour quality, while the wage in dredging is considerably her than ın manufacturing

[The second aspect is the extent to which export industries develop a ıned labour force which subsequently can be employed elsewhere ın › economy, thus generating unpriced externalities for other industries ble VI shows that in terms of skilled workers per unit of output, there ıttle difference between rubber estates and the tin industry, since the , proportion of skilled workers ın the estate labour force ıs counter-lanced by rubber's very high over-all ratio of labour to output. The ıeration of skilled workers per unit of investment is shown to be similar rubber and dredging (again, because rubber's low skill requirements counterbalanced by high over-all labour requirements) The skilled rker.investment coefficient is much higher for gravel-pump mines, but should be noted that capital equipment on gravel-pump mines has a ch shorter life than is the case in dredging or rubber [Thoburn, 1971,

pp. 501–2]. It is important to add that the type of skilled labour employed on rubber estates and more particularly on mines consists of fitters, boiler-makers, electrical chargemen, etc , all of whose skills are highly usable by local manufacturing industry. These remarks apply with even greater force to the generation of a force of skilled workers by the local engineering industry, which has developed to supply export industries' capital goods requirements [Thoburn, 1972*a*, pp 21–2] The development of a skilled labour force is helped by the fact that in both sectors of the tin industry, in engineering, and to a lesser extent in rubber planting, workers are trained through apprenticeship schemes These transfer the cost of training from the employer to the worker, and lessen the disadvantage to any one firm or industry of workers leaving to take employment elsewhere This means of course that labour training is not strictly an 'externality' provided by export sector *firms*. Nevertheless, the economy benefits from this labour training which results from the existence of the export industries.

VII. Conclusions

The study has concentrated on establishing the proportion of export income initially retained in Malaysia [1] It has been shown that in both foreign sectors—tin dredging and rubber estates—over 70 per cent of export income is initially retained The payments structure of these two sectors is quite different, however Dredging has a very high proportion of pre-tax profits in its payments, with a small wage bill, and a small but not insignificant set of materials purchases The channelling of dredging income back into the domestic economy has been to a great extent the result of government tax policy. This policy is largely a postwar phenomenon, although the export duty on tin is over a hundred years old and was used to finance railway construction and other government development expenditures in the early days Also, the purchase by local inhabitants of dredging shares on a large scale has meant that dividend payments too have been channelled back into Malaysia Gravel-pump mining, the local tin sector, also has high payments retention, but with a larger proportionate wage bill and materials purchases. In this case the distribution of export income between domestic and foreign recipients is little affected by the distribution between labour, profit receivers, and the Malaysian government.

Rubber estates, in contrast to tin mines, have high proportions of wages in total output value. Profits are larger than might be expected from this

[1] Of course, export income must eventually flow out of the country (or be retained as foreign exchange reserves), but it is its multiplier-accelerator effects in the meantime which are important.

fact because materials purchases are slight (in proportion to the valu output) Here, too, government tax policy has channelled profits back Malaysia There has been little acquisition of foreign rubber comp shares by local people, but the local ownership of estate land is substan and there is complete local ownership of smallholdings. Malays in p ticular have participated in the rubber industry as smallholders

A high proportion of purchases made by export sector workers is of loc produce, though in the rubber industry there is little relation, at least the short run, between changes in export earnings and changes in the wa bill [1] The indirect import content of local production and the import conte of expenditures of local profit earners and executives, and the governme out of income accruing from export earnings, were not examined becau of lack of data (and research time) To the extent that secondary outflow occur from these sources, the findings of the study are weakened.

The tin industry, and rubber to a lesser extent, has generated importa linkages related to its size, and there is some evidence to suggest (i contradiction to the Little–Mirrlees position discussed in Section I) th the linkage investment has not been at the expense of other investmen What has made for this high-linkage situation in tin? First, the maj material input is power—first coal, then diesel and electricity Coal h high transport costs relative to its value, while electricity (with th exception of some minor past imports from Singapore) is a non-trade good Hence the usual expedient of importing intermediate products w limited Engineering, the other major linkage of tin (and rubber), is a industry whose products are highly specialized and divisible Thus it wa possible, as I have shown elsewhere [Thoburn, 1972a], for engineering t begin with repair and maintenance, and progress to simple erection wor and finally to fabrication of quite complex machinery while importin those components difficult to produce locally

One particularly important effect of the development of an expo industry is on the quality and earnings of the local labour price employe in the industry Here both gravel pumping and dredging have been a effective in improving labour quality as the existing manufacturing indu tries in Malaysia This is especially important in gravel pumping, whos over-all ratio of labour to gross output is substantially greater tha dredging or manufacturing Rubber estate earnings are very much lowe even though the industry is unionized (as is dredging, but not grave

[1] Thus the impact of rubber exports on the domestic economy does not seem to hav worked through the multiplier effects of changes in export earnings, but through the exis ence of a large 'autonomous' component in domestic expenditure out of export earning On the acceleration side, the large and discontinuous nature of investment in electricit supply, petroleum refining, or chemical fertilizer production does not make it meaningful t attempt to calculate an accelerator of the sort used in macro-models

pumping). The income of rubber smallholders has not been investigat in detail for this study, but existing studies indicate that it is actually ve low in many cases,[1] but potentially high on adequate sized plots plant with high-yielding strains of rubber.[2]

It has been shown that the tin-mining industry, especially the loc sector, has promoted development in most ways open to an expor industry, though the channelling of dredging profits to the Malaysian government is relatively recent [3] Rubber has not improved the labour quality greatly, nor generated large linkages, but its enormous foreign exchange earnings have enabled Malaysia largely to avoid the balance of payments constraint faced by many less developed countries. At least its high retention of export earnings destroys the simple view of the industry as an enclave, and its high labour requirements will help Malaysia to cope with the unemployment problems created by the country's rapid population growth, so long as supplies of undeveloped land remain to permit the industry's further growth and external market conditions remain favourable Moreover, many other important effects of export growth do not lend themselves to the kind of quantification used here [4] Rubber has spread the habit of cash crop cultivation over almost the whole of West Malaysia It has provided an important regular source of income to rural Malay households, helping to reduce the danger of indebtedness associated with income from annual crops [Swift, 1965, p. 74] The trading and manufacturing opportunities associated with tin and rubber (and other exports), and the savings generated from profits from exports, have stimulated the growth of local entrepreneurship and responsiveness to economic opportunity The transport system and urbanization associated originally with tin mining not only facilitated the development of rubber but also helped widen the domestic market over most of the west coast of the country

University of East Anglia, Norwich

[1] Fisk (1961) estimates that income per worker per month in a Malay area of Selangor is only $31

[2] Barlow and Chan (1968) estimate an income of $164 a month for a family holding of 6 8 acres planted with high-yielding material at a rubber price of 63 Malaysian cents per lb With high yielding rubber tapped alternate-daily, a family with two working members could well manage a holding of at least twice that size. This potential could be realized to the extent that the large amount of unused land suitable for agriculture in Malaysia is cleared and brought under cultivation Under the Second Malaysia Plan, 1971-5 (Prime Minister's Department, 1971) it is planned to increase the rate of land development to 200,000 acres per year, a figure equal to about 12 per cent of the existing rubber acreage Of course rubber planting is not the only means of raising peasant income Oil palm is an important alternative, but further discussion of policy is outside the scope of this paper

[3] It is likely that tin's generation of final demand linkages may be more important than it first seems. The over-all results are dominated by rubber industry workers' purchases Tin workers with higher incomes would demand more sophisticated products

[4] These can be seen as unpriced externalities of export growth.

REFERENCES

BALDWIN, R. E. (1963), 'Export technology and development from a subsisten level', *Economic Journal*, vol. lxxiii, No. 289.

—— (1966), *Economic Development and Export Growth. A Study of Northern Rhodesi 1920–66*, Berkeley, California University Press.

BARLOW, C., and CHAN CHEE KHEONG (1968), 'Towards an optimum size of rubb holding', Natural Rubber Conference preprint, Kuala Lumpur.

BAUER, P. T. (1948), *The Rubber Industry. A Study in Competition and Monopo* London, Longman, Green & Company.

Central Electricity Board, Federation of Malaya/Malaysia (1950 and 1967), Annu Reports, Kuala Lumpur.

DRABBLE, J. H. (1967), 'The plantation rubber industry in Malaya up to 1922 *Journal of the Malaysian Branch of the Royal Asiatic Society*, vol. xl, Pt. 1.

Electricity Department, Federated Malay States (1937), *Annual Report*, Kuala Lumpur.

FISK, E. K. (1961), 'Productivity and income from rubber on an established Malay reservation', *Malayan Economic Review*, vol. vi, No. 1.

HARVIE, C H. (1964), 'Export multipliers and the stability of the Federation of Malaya's economy', *Malayan Economic Review*, vol. ix, No. 1.

HIRSCHMAN, A O. (1958), *The Strategy of Economic Development*, New Haven, Yale University Press.

LEVIN, J. V. (1960), *The Export Economies. Their Pattern of Development in Historical Perspective*, Cambridge, Mass , Harvard University Press.

LIM CHONG YAH (1967), *Economic Development of Modern Malaya*, Kuala Lumpur, Oxford University Press.

LIM, Y. (1968), 'Trade and growth: the case of Ceylon', *Economic Development and Cultural Change*, No 2

LITTLE, I. M. D., and MIRRLEES, J. A (1968), *Manual of Industrial Project Appraisal in Developing Countries*, vol. ii, *Social Cost-Benefit Analysis*, Paris, Organization for Economic Cooperation and Development.

—— —— (1972), 'A reply to some criticisms of the OECD Manual', *Bulletin of the Oxford University Institute of Economics and Statistics*, vol 34, No 1

MEIER, G M. (1968), *The International Economics of Development. Theory and Policy*, New York, Harper & Row

Mines Department, Federation of Malaya/Malaysia (1955 and 1968), *Bulletin of Statistics Relating to the Mining Industry of Malaya/Malaysia*, Kuala Lumpur.

Prime Minister's Department, Malaysia (1971), *Second Malaysia Plan*, 1971–5, Kuala Lumpur.

Railways Department, Federated Malay States (1916), *Annual Report*, Kuala Lumpur.

—— (1935), *Fifty Years of Railways in Malaya 1885–1935*, Kuala Lumpur.

ROSENBURG, N. (1971), *The Economics of Technological Change, Selected Readings*, London, Penguin Books.

Rubber Research Institute of Malaya (1969), *Report on the 1964 Survey of Estates*, Kuala Lumpur.

SCITOVSKY, T. (1954), *Papers on Welfare and Growth*, London, Allen & Unwin.

SINGER, H. W. (1950), 'The distribution of gains between investing and borrowing countries', *American Economic Review, Papers and Proceedings*, vol. ii, No. 2.

SOLO, R. (1966), 'The capacity to assimilate an advanced technology', *American Economic Review, Papers and Proceedings*, vol lvi, No. 2 reprinted in Rosenburg (1971) (to which page numbers refer).

Statistics Department, Federation of Malaya/Malaysia (1968a), *Census of Manufacturing Industries*, Kuala Lumpur.

—— (1968*b*), *Census of Mining Industries*, Kuala Lumpur.
—— (1965), *Interindustry Accounts*, Kuala Lumpur.
—— (1970), *Monthly Bulletin of Statistics*, Kuala Lumpur.
—— (1952, 1966, 1968*c*), *Rubber Statistics Handbook*, Kuala Lumpur

STEWART, F., and STREETEN, P. P. (1972), 'Little–Mirrlees methods and project appraisal', *Bulletin of the Oxford University Institute of Economics and Statistics*, vol. 34, No. 1.

SWIFT, M. G. (1965), *Malay Peasant Society in Jelebu*, London, The Athlone Press.

THOBURN, J. T. (1971), 'Exports in the economic development of West Malaysia', Edmonton, Canada, University of Alberta Ph.D. thesis.

—— (1972*a*), 'Exports and the Malaysian engineering industry. A case study of backward linkage', *University of East Anglia, Norwich, Economics Discussion Paper No. 5*

—— (1972*b*), 'Ownership of shares in UK-incorporated public rubber planting and tin dredging companies in Malaysia', *Kajian Ekonomi Malaysia*, vol VII, No 2

—— (1972*c*), *Theories of Trade and Growth in Less Developed Countries*, unpublished monograph, University of East Anglia, Norwich.

WATKINS, M. H. (1963), 'A staple theory of economic growth', *Canadian Journal of Economics and Political Science*, vol XXIX, No. 2.

WOLF, C, and SUFRIN, S. C. (1955), *Capital Formation and Foreign Investment in Underdeveloped Areas*, Syracuse, Syracuse University Press.

YIP YAT HOONG (1968), 'Recent changes in the ownership and control of locally incorporated tin dredging companies in Malaya', *Malayan Economic Review*, vol. 13, No. 1.

ZORN and LEIGH-HUNT, Messrs. (1969*a*), *Manual of Rubber Planting Companies*, London, privately published.

—— (1969*b*), *Manual of Tin Mining Companies*, London, privately published.

DISUTILITY OF EFFORT, MIGRATION, AND THE SHADOW WAGE-RATE[1]

By DEEPAK LAL

Introduction

THE problem of socially valuing labour (that is, the shadow wage-rate (*SWR*)) in economies subject to population pressure, has been widely discussed in the economic development literature [**2, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 18, 19, 21, 22, 24, 26, 31, 35, 38**]. The *SWR* is defined as that magnitude to which the marginal productivity of labour should be equated to maximize feasible social welfare[2] Though this might appear to be a question in pure positive economics, it has become an area where explicitly or implicitly strong normative disagreements have crept in Some of these disagreements are about 'values', in respect of the components of the social welfare function, whilst others are about the relevance of policy constraints and aspects of the 'factual' environment to the determination of the *SWR* The first few sections of the paper are devoted to incorporating, in a single derivation of the *SWR*, the various elements emphasized by different writers, whilst the last section will point out the limits, in my view, of the significance of many of these aspects for determining the *SWR* in practice, in most developing countries.

In past discussions two sectors have usually been postulated –a small, high-productivity sector (industry), which draws in labour from a large, low-productivity sector (agriculture). The correct shadow wage in the high-productivity sector is then the subject of debate, in which two major concerns can be identified First, whether or not the existence of surplus labour (in the sense that an increase in industrial employment does not lead to a fall in output elsewhere in the economy) theoretically requires the marginal product of a labourer in the low-productivity sector to be zero, and, moreover, whether empirically this latter assumption is justified [**2, 11, 14, 27, 29, 30, 31, 32, 40**]. Secondly, even if the output forgone elsewhere in the economy is zero, whether in a non-optimal savings situation,[3] it would be correct to value newly employed labour as costless, when increased employment leads to an increase in aggregate current consump-

[1] I have benefited from discussions with Professors A C Harberger, I M D Little, and A K Sen, and Messrs. D Mazumdar and M Fg Scott, and from the comments of members of seminars at the Delhi School of Economics and London School of Economics. The research on which the first draft of this paper was based was done whilst I was a Research Fellow of Nuffield College, Oxford

[2] This definition should be distinguished from the shadow price of labour in the programming sense See Sen [**31**]

[3] In terms of the standard two-period Fisherian capital-theoretic model, savings non-optimality implies that the rate of transformation between present and future consumption is greater than the rate of indifferent substitution

tion, which *ex hypothesi* is not as valuable as current savings [**4, 5, 18, 21, 22, 24, 31, 35**].

There is an emerging consensus that both output forgone and the costs of the increased consumption resulting from extra employment must be taken into account in determining the shadow wage-rate [**22, 24, 31**] Most analyses, however, have not considered the implications for the shadow wage-rate of incorporating changes in the disutility of effort, and the rural–urban migration, which will normally accompany an increase in industrial employment The analyses which have recognized the importance of the leisure-income choice, in determining the supply of labour to the industrial sector [**2, 32**], have concentrated on the implications for output changes in the sector from which labour is withdrawn, but the ensuing implications for the social valuation of labour have not been drawn, whilst the analyses which have considered rural–urban migration [**37, 38**] have not drawn the full implications for the *SWR* entailed by their models They have also made restrictive assumptions about agricultural output changes which follow a withdrawal of workers from that sector, and in the particular form of the migration function Finally, there has been an implicit assumption, in most discussions of the *SWR*, that no social value should be attached to the changes in private disutilities of effort *per se*, which an increase in employment may entail, or else that the private marginal disutility of labour is zero over the relevant utility range

Section I presents the traditional formulation of the *SWR* Section II examines the changes which have to be made in this traditional formulation, when the effects of the income-leisure choice, and rural–urban migration on output forgone, are incorporated Section III derives the changes in the *SWR* formulation necessitated by incorporating the costs of the disutility of effort which may accompany employment changes, whilst Section IV presents arguments why the social valuation of these disutilities should perhaps be zero as well as for the limited significance of other factors which have been considered relevant in determining the *SWR* Table I summarizes the notation used in the paper

I

The traditional *SWR*

The particular formulation of the *SWR* I shall consider is that due to Little–Mirrlees [**22**], but except for a change in numeraire (which they take to be savings rather than consumption) their analysis is similar to the other well-known ones due to Sen [**31**] and Marglin [**24**], and hence the remarks in this paper are relevant to all these existing formulations of the *SWR*

Traditionally the *SWR* is derived as follows Assume first that the wage

TABLE I

Summary of notation

a	income per worker in agriculture.
c	migration costs per worker.
D	the savings equivalent of the net change in the disutility of effort caused by employing one more man.
L	the supply price of rural–urban migrants ($= c+a$).
m	the marginal product of labour in agriculture.
M	number of migrants when one more 'organized' urban sector job is created.
$(P_e)\,P$	the (equilibrium) probability of finding a job at the industrial wage.
s	the premium on savings *vis-à-vis* consumption.
SWR	the shadow wage-rate.
U_t	the number of urban 'unemployed' seeking organized urban sector jobs in period t.
V_t	number of industrial job vacancies in period t.
W_i	the industrial (market) wage in the 'organized' urban sector.
W_u	average earnings per worker in the 'unorganized' urban sector.
y	the change in agricultural output when one worker is withdrawn.
λ	the social value placed on the savings equivalent of the private disutilities of effort.

paid to a labourer in his new job, W_i, is above the output forgone elsewhere by moving him from his previous employment, y. Furthermore, assuming that workers in both industry and agriculture consume all their incomes, there is the cost of the extra consumption (W_i-y) the economy is committed to, as $W_i > y$. But not the whole of this increase in consumption is a social cost, as society does value s units of consumption as equal in social value to one unit of savings.[1] Hence, from the total increase in consumption, a proportion $(1/s)$ must be subtracted to get the net social cost (in terms of the numeraire, savings) of the increased consumption. Hence the social cost of employing one more person will be

$$SWR = W_i-(W_i-y)/s. \tag{1}$$

II

Output forgone and the social cost of extra consumption

The leisure-income choice

This section examines the validity of the traditional identification of the output forgone, y, in the SWR formulation, with the marginal product of labour in agriculture, m. Whilst this assumption is valid for landless agricultural workers in agriculture, it will not hold, in general, for family farm workers on farms without any hired labour. Moreover, in most traditional analyses it was also assumed that, in a surplus labour economy,

[1] The value of 's' has to be determined within a Ramsey-type variational formulation of intertemporal optimality. [See **3, 4, 18, 28, 34, 35**.]

the marginal product of the labourer withdrawn from the traditional sector would be zero ($m = 0$), [**19, 21, 24, 26, 31**], and hence $y = 0$

In a definitive analysis of dualism and surplus labour, using a model of family farms on which there is equal work and income sharing, and which explicitly incorporates leisure as an argument in the individual peasant's utility function, Sen [**32**] has shown that zero marginal productivity is not a necessary condition for the existence of surplus labour The necessary and sufficient conditions are being given by a constant disutility of effort, which implies a constant marginal rate of substitution between income and leisure, over the relevant range of hours worked per man, in the traditional sector.[1] Thus positive marginal productivity and surplus labour are compatible. For the marginal productivity of agricultural family farm workers to be zero, they would have to be satiated with leisure

Furthermore, from this analysis it is evident that the y term in the *SWR* formulation cannot be taken to be the marginal product of the withdrawn worker in agriculture, as the change in output, following the withdrawal of a worker from agriculture, will not in general be equal to his marginal product if the worker is withdrawn from a family farm without hired labour. For instance, the marginal product may be positive and yet there may be no change in output with an agricultural worker's withdrawal if the disutility of effort for family farm workers is constant over the relevant utility range For the output forgone to equal the marginal product of the worker withdrawn from a family farm, each worker would have to work an invariant number of hours both before and after the withdrawal of the family labourer.[2]

Divergence between average and marginal wage cost

Moreover, Dixit [**5**] has recently emphasized that the traditional analysis may *understate* the extra consumption cost of industrial labour. This is due to the assumption made in these analyses that 'agricultural' workers can be hired by the 'industrial' sector at a constant real wage (W_i), which is either given by a constant institutional wage, or else by a constant supply price of labour to the industrial sector Dixit suggests that this assumption

[1] Also see Berry and Soligo [**2**], and Stiglitz [**36**]

[2] In general the elasticity of output with respect to the number of working members E will be given by the following expression

$$E = G \frac{n+u}{n+u \quad G+g} \quad \text{(see Sen [32] for the derivation)},$$

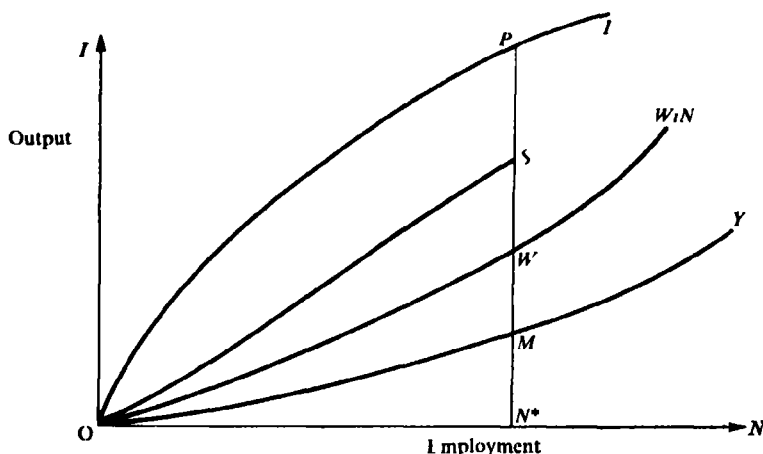
where G is the elasticity of output with respect to hours of labour, n the elasticity of marginal disutility of work with respect to the hours worked per worker, u is the elasticity of the marginal utility of income with respect to per worker income, and g is the elasticity of the marginal product of labour with respect to the total number of man hours worked In the special case of surplus labour, both n and u are zero, and hence $E = 0$ For the change in output on one worker's withdrawal to be equal to the marginal product of labour, E would have to equal G, which could happen if 'n' were very large

may be unrealistic, especially if there are terms of trade effects following a withdrawal of labour from agriculture Then, if the industrial labour market is competitive, the supply price of labour to industry and, hence, the industrial wage will rise with increased industrial employment.[1] This will create a divergence between the *average* (W_i) and *marginal* ($W_i+\Delta W_i$) cost of hiring industrial labour The extra consumption the economy will be committed to will then be given by the difference between the *marginal* cost of hiring ($W_i+\Delta W_i$) and output forgone y [2] Hence

$$SWR = (W_i+\Delta W_i)-(W_i+\Delta W_i-y)/s, \tag{2}$$

and if the premium placed on savings is very high ($s \to \infty$), the SWR will

[1] In the simple closed economy two-sector model analysed by Dixit, the supply price of industrial labour is equal to the income forgone by agricultural family workers moving to industrial jobs In short-run equilibrium their income forgone is determined by the average physical product of labour in agriculture (assuming equal income-sharing amongst family farm workers) and the relative price of agricultural output With the withdrawal of an agricultural worker, the average product of labour in agriculture rises, whilst total agricultural output (assuming no surplus labour) falls This last factor leads to a rise in the relative price of agricultural output The net effect is to raise the average value product of labour in agriculture, and hence the supply price of labour to the industrial sector.

[2] This can be seen heuristically from the accompanying diagram OI is the industrial production function, with a given fixed capital stock and variable employment (N) The OY curve gives the total output forgone, and the OW_iN curve the total wage bill for each level of industrial employment The shape of both these curves reflects the assumed rising output forgone and wage rate (= agricultural income forgone) of industrial labour Assume that there is an infinite premium on savings, and all wages are consumed Social welfare is then maximized with the industrial employment level ON^*, where the marginal wage cost (slope of the tangent at W) is equal to the marginal product of labour (slope of the tangent at P) Hence to ensure the optimal level of industrial employment, a wage *tax* of WS/ON^* will have to be imposed (The slope of OS being equal to the slope of the tangent at W) The SWR is then given by $SN^*/ON^* = WN^*/ON^* + WS/ON^* \quad W_i + \Delta W_i$

It is then easy to determine, that when $s > 1$, but not infinite, the SWR will be given by expression (2) in the text

be higher than the market wage (W_i)[1] Note, however, that if there is a constant institutional wage in the industrial sector, then $\Delta W_i = 0$, and the SWR will be given as before by (1)

Rural–urban migration

Furthermore, as certain models of labour markets in developing countries have emphasized [37, 38], the impact on net output in the economy cannot be deduced from the impact effects on output in the sector from which the new worker may be withdrawn, and to obtain the value of y it will be necessary to trace through all the indirect effects in terms of the rural urban migration that may ensue as the result of creating one more job in the urban sector Suppose M people migrate from rural areas, and the resulting change in agricultural output is My Then,[2] the

$$SWR = W_i - (W_i - My)/s \tag{3}$$

We next need to specify the determinants of M One particular model of rural–urban migration due to Harris and Todaro [38] has recently gained wide currency In this model Harris and Todaro assume that there is *no* surplus labour in agriculture Agricultural workers receive their marginal product, m They migrate to the cities because the expected income in the urban sector is just equal to the income they forgo in agriculture. The expected urban income is determined by the probability (P) of finding urban employment at the industrial wage W_i At the margin, therefore, migrant workers will equate their incomes (= marginal product) in agriculture, m, to the expected urban wage, PW_i (That is $m = PW_i$)

P is given, in the Harris-Todaro model, by the ratio of the employed to the total urban labour force[3] This seems unrealistic A more plausible determinant of the chances of a single migrant is given by the number of vacancies (V_t) occurring per period (t) divided by the number of candidates for these vacancies, that is the urban unemployed (U_t)[4] However, on

[1] ΔW_i can be computed directly from fig, p 116 n 1 above as follows When $s \to \infty$, the SWR is $\frac{d(W_i N)}{dN} = W_i + \frac{N dW_i}{dN}$. Hence $\Delta W_i = \frac{N dW_i}{dN}$. To determine $\frac{dW_i}{dN}$, the production and demand relationships in the two sectors need to be specified Dixit [5] shows that, in his simple two-good two-sector model, the demand relationships are more crucial determinants of dW_i/dN than the production relationships

Furthermore, it may be noted that given the rising output forgone curve OY, the *average* output forgone per industrial worker, will also be less than the *marginal* output forgone However, in our discussion in the previous section, we have already taken this into account in the determination of the y term in the SWR formulation In terms of Fig 1,

$$y = Y/N + N\ dY/dN$$

and its determinants are summarized by the expression given in p 115 n 2

[2] Remembering that if there is a divergence between average and marginal costs, the W and y terms will be the relevant *marginal* values This proviso applies to all the formulae containing W_i and y in the remainder of the text

[3] See [38], p 128 This is also the assumption made by Harberger [10], p 570

[4] The latter in fact was the determinant of P in the earlier Todaro formulation. (See [37], p 142)

either formulation of P, *its equilibrium value* (P_e) will be determined by the equilibrium migration condition $m = P_e W_i$, that is $P_e = m/W_i$. Hence, P_e depends solely on the rural–urban income differential.

Now, suppose in period (t), $P_t = \frac{V_t}{U_t} = P_e = \frac{m}{W_i}$. Then there will be no rural–urban migration in this period. However, in the next period ($t+1$) the number of job-seekers left in the urban labour market will be $(U_t - V_t)$, and if the number of vacancies is $V_{(t+1)}$ ($\geqslant V_t$). Then, $P_{t+1} = \frac{V_{(t+1)}}{U_t - V_t} > P_t = P_e$, and this will induce M_{t+1} rural migrants as job-seekers, till,

$$P_{t+1} = \frac{V_{(t+1)}}{(U_t - V_t) + M_{t+1}} = P_e = \frac{m}{Wi}.$$

From this it is easy to derive that $M_{t+1} = \frac{W_i}{m}(V_{t+1} - V_t) + V_t$. From this it is obvious that, if the number of vacancies remains constant ($V_{t+1} = V_t = V$), then M in each period will equal V. If between two periods the number of vacancies increases by dV[1] ($= V_{t+1} - V_t$), then the increase in migration dM ($= M_{t+1} - M_t$) will be $dM = (W_i/m)dV$ Thus if one extra job is created in the industrial sector this will induce the migration of

$$M = W_i/m = 1/P_e$$

individuals from the rural sector Substituting this value of M in equation (3) above and remembering that in the Harris–Todaro model $y = m$, we get the $SWR = W_i$, the industrial wage.[2,3]

However, as noted above, in general it cannot be assumed that the change

[1] The number of vacancies may rise because total industrial jobs are increasing and/or because of a higher rate of turnover of industrial jobs In the latter case, assuming that the industrial workers losing their jobs begin to search for new industrial sector jobs, the increase in industrial vacancies due to a higher turnover rate will be matched by an equivalent increase in the number of industrial job-seekers Therefore, there is no *net* effect on the probability of finding an industrial sector job, and hence no effect on migration Though naturally the *unemployment rate* in the urban sector will be affected Thus for our purpose dV should reflect the *net* increase in industrial vacancies, and as such labour turnover in the industrial sector will not affect the number of migrants to the town For a model which includes the rate of industrial labour turnover as a determinant of the urban unemployment rate in a migration model, see G E Johnson [12]

[2] I owe this point to Al Harberger (see [10]), who has further extended the model to the case where the migrating peasants do not equate their income forgone in agriculture (which is assumed to be equal to their marginal product m) to the expected income in the towns PW_i, but are risk-averters So that in equilibrium, that is with no rural–urban migration, $PW_i > m$ Harberger argues that, in this case, if in our measure of social welfare we respect individual preferences, the industrial wage W_i will still be the correct shadow wage, as the migrants may be considered to regard the probability P of income W_i to be equivalent to the certainty of income m This, however, assumes that social welfare should be based on expected (*ex ante*) utility, rather than on realized (*ex post*) utility

[3] It may also be noted that the equilibrium migration condition $P_e = m/W_i$, simultaneously ensures equilibrium on both the *stock* aspect of the numbers employed to the total equilibrium level of the urban labour force, and on the *flow* aspect of the equilibrium flow of migration generated by a given rate of change in urban employment.

in output in the agricultural sector (y) will equal the marginal product of labour (m). In that case, the change in output within the Harris–Todaro migration model will be given by y/P_e, where y is the change in output in agriculture when one worker is withdrawn. The equilibrium migration condition will be $P_e = a/W_i$ (where a is the income the worker received in agriculture, which on a family farm would be equal to the average product of the farm if we assume equal income and work sharing on family farms). Hence the total output forgone will be $y.W_i/a$, and the SWR by substitution in expression (1) or (3) will be

$$SWR = W_i - W_i(1-y/a)/s$$

From this it is obvious that on the special Harris–Todaro assumption that $y = m = a$ the SWR will equal W_i the industrial wage. These implications of their model were not drawn by Harris–Todaro.

More important, however, the Harris–Todaro migration model is also restrictive in other respects. First they consider the migration decision as a one-period decision, whereas strictly it should be a multi-period decision in which the present value of the costs of migration should at the margin be equal to the present value of the benefits from migration.[1] If, however, as seems likely, most migrants have a fairly high subjective rate of time preference (fairly short time horizon), then the use of a single-period migration decision function may not be invalid. Secondly, Harris–Todaro do not incorporate any of the costs of migration (real and/or 'physic'),[2] nor the relatively higher costs of urban living which the migrant would have to incur in their migration function Finally, and most important, their migration model fails to take account of the existence of a fairly competitive 'unorganized' (services and small industry) sector urban labour market with high labour turnover and easy entry for new workers, which is typical of many developing countries, and which provides some income to the migrants whilst they are searching for an 'organized' (industrial) sector job at the high institutional wage W_i.[3]

Thus it is essentially the last two features which need to be incorporated into a more general migration function. To derive the SWR for this more general migration model, we continue to assume that a one-period decision model is a fair approximation to reality.[4] However, we now assume that in addition to the agricultural income forgone, a, the migrant has to incur migration costs, which include both the real and 'phychic' costs of migrating, as well as the relatively higher costs of urban living to maintain the

[1] Todaro [37], p 143 n 10, notes this.

[2] Though Harris–Todaro note the existence of these costs, see [38], p 129 n 8

[3] This has been noted and stressed by Mazumdar [23]

[4] The one-period general migration model presented in the text can easily be extended to the multi-period case See Mazumdar [23].

same standard of living he enjoyed in the countryside. Let the sum of both these costs of migration be c. The sum of income forgone a, and the migration costs c, will then be the supply of price of labour L $(= a+c)$ Finally, we assume that if the migrant does not succeed in obtaining an industrial sector job at the high institutional wage of W_i, he will nevertheless find some employment in the 'unorganized' urban labour market and derive an income per period of W_u Given that the probability of getting an 'organized' (industrial) sector job is P, then at the margin the migrant will equate the total costs of migration, which are given by L, with the expected benefits $(PW_i+(1-P)W_u)$, that is in equilibrium

$$L = PW_i+(1-P)W_u$$

and the equilibrium value of $P_e = (L-W_u)/(W_i-W_u)$

As before, with the creation of an extra industrial sector job $M = 1/P_e$ migrants will move from agriculture, and as the output forgone per migrant in agriculture is y, the total output forgone will be

$$My = y(W_i-W_u)/(L-W_u),$$

and the

$$SWR = W_i-[W_i-y(W_i-W_u)/(L-W_u)]/s. \tag{4}$$

III

Disutility of effort

Having examined the effects of the income-leisure choice and rural–urban migration on changes in output forgone with increased industrial employment and the consequent modifications of the traditional SWR, we now turn to examine the implications of recognizing that, in addition to changes in output, there will also be a net change in the aggregate disutility of effort with increased employment λD will be added to the right-hand side of expression (1), where D is the savings equivalent of the net change in the disutility of effort caused by employing one more man, and λ is the value society places on this cost

We can furthermore derive the consumption equivalent of D. Assume initially that there are no imperfections in the labour market Then, at the margin, utility maximizing workers will equate the disutility of increased effort with the utility from the increased incomes (which we assume are all consumed) this extra work makes possible That is the extra disutility must equal the change in workers' consumption (including that of those left on the farm) which is given by (W_i-y)—the difference between the industrial wage and the total output forgone by employing one more man in the industrial sector The value of this consumption equivalent of the net change in the disutility of effort in terms of savings (our numeraire) will be $D = (W_i-y)/s$ and the

$$SWR = W_i-(W_i-y)(1-\lambda)/s.$$

Next relax the assumption that all labour markets are competitive, and assume that there is an institutional wage in the sector to which the labour is moving which is above the supply price of labour L. The latter term includes all the private disutilities that may attach to the new job. Our earlier expression for the consumption equivalent of the change in disutility $(W_i - y)$ will now be overstating the true change in disutility by $(W_i - L)$, which is the difference between the institutional wage W_i and the supply price of labour L. The net change in disutilities in this more general case will, therefore, be given by $(W_i - y) - (W_i - L) = (L - y)$, and as before, the value in terms of savings will be $D = (L-y)/S$, and the

$$SWR = W_i - [W_i - y - \lambda(L-y)]/s \qquad (5)$$

If $\lambda = 0$, that is society places no value on the change in the private disutilities of effort, we get the traditional SWR as in (1) above.[1] If, however, it is assumed that society should value disutilities of effort at their private cost, then $\lambda = 1$, and the

$$SWR = W_i - (W_i - L)/s \quad \text{or equivalently}$$

$$= L + (W_i - L)(1 - 1/s).$$

The first term is the supply price of labour, the second is the value in terms of savings of the extra consumption generated by the excess of the institutional wage over the supply price of labour. Thus when $\lambda = 1$, we get the standard neoclassical result that the SWR will be the supply price of labour if there is no divergence between the social value of present consumption and savings, that is $s = 1$, and furthermore, that if $W_i = L$, that is if labour markets are competitive, the SWR will equal the market wage W_i, no matter what the value of s, and irrespective of any divergence between y (the output forgone elsewhere in the economy) and the industrial wage W_i.[2]

We next introduce rural–urban migration. First, *à la* Harris–Todaro. We know from the above analysis that, in the general case, the net change in disutilities is given by $(L-y)$, where L is the supply price of labour and y is the net change in output elsewhere in the economy resulting from increased employment. From our analysis in Section II we know that in

[1] This would also be the case if the value of D was zero, that is if there was leisure satiation. But then that would imply that $y = 0$, and the marginal product of labour in agriculture would also be zero ($m = 0$).

[2] The assumptions underlying a number of well-known discussions of the real cost of labour in a surplus labour economy can be seen from expression 5. (I) Galenson–Leibenstein [9] and Dobb's [6] SWR = market wage W_i, implies that either (i) $W_i = y$ and $\lambda = 0$ or (ii) $y = 0$, $s \rightarrow \infty$, and $\lambda = 0$, or $D = 0$, (II) for Kahn [15], Lewis [19], the $SWR = 0$ which implies that $y = 0$, $s = 1$, and $\lambda = 0$ or $D = 0$, (III) for Sen [31] and Marglin [24], the $SWR = W_i(1 - 1/s)$, which implies that $y = 0$, and $\lambda = 0$ or $D = 0$, (IV) for Little–Mirrlees [22] the $SWR = W_i - (W_i - y)/s$, which implies $\lambda = 0$. As they assume a positive marginal product in agriculture D cannot be zero, (V) for Harberger, [10], the $SWR = L$, which implies (i) $\lambda = 1$, $W_i = L$, or (ii) $s = 1$, $\lambda = 1$.

the Harris–Todaro model, total output forgone is $y.W_i/a$. We also know that as with the employment of one extra industrial worker

$$M = 1/P_e = W_i/a$$

workers will be drawn from agriculture and as each worker's supply price is a, their total supply price will be $M.a = W_i$. Hence the net change in disutilities will be $W_i(1-y/a)$ The social value of the cost of this increased disutility of effort, in terms of savings, is $\lambda W_i(1-y/a)/s$, and hence the SWR in the Harris–Todaro model when disutilities of effort are taken into account is

$$\begin{aligned} SWR &= W_i - W_i(1-y/a)/s + \lambda W_i(1-y/a)/s \\ &= W_i - W_i(1-y/a)(1-\lambda)/s. \end{aligned} \tag{6}$$

Once again, if we make the particular Harris–Todaro assumption that $y = a$, that is the change in agricultural output when one peasant is withdrawn is equal to his income in agriculture, the $SWR = W_i$; the industrial wage irrespective of the values of s and λ. Whilst if $\lambda = 1$, then irrespective of the value of s and any divergence between y and a, the industrial wage is again the SWR.

Next consider the more general migration model of Section II The supply price of each migrant in this model was $L = a+c$, and with every extra job created in the organized industrial sector we had

$$M = 1/P_e = (W_i - W_u)/(L - W_u)$$

migrants from rural areas. Their total supply price is therefore

$$M.L = L(W_i - W_u)/(L - W_u).$$

As before, the total output forgone[1] is $y(W_i - W_u)/(L - W_u)$, and hence the net change in disutilities (which is given by the difference between the total supply price of labour and output forgone) is

$$(L-y)(W_i - W_u)/(L - W_u).$$

Hence, by substitution in expression (5) we have

$$SWR = W_i - (1/s)[W_i - \{y + \lambda(L-y)\}(W_i - W_u)/(L - W_u)]. \tag{7}$$

If no social value is placed on the private disutilities of effort, $\lambda = 0$, and the above expression reduces to expression (4) above.

If, however, the private disutilities are socially valued at par $\lambda = 1$, and the

$$SWR = W_i - (1/s)[W_i - L(W_i - W_u)/(L - W_u)].$$

Thus in this more general, and more realistic, migration model, none of the simpler SWR derivations from the Harris–Todaro [38] and Harberger [10] type models will hold.

[1] It should be noted that in this case where it is assumed that there are 'real' resource costs involved in migration (part of the component of L), these costs must be added to the output forgone term y

IV

Normative parameters

There are two policy parameters in the above formulae —s, the premium on savings *vis-à-vis* consumption, and the social valuation of the disutility of effort λ The existing literature has discussed the reasons for valuing $s > 1$ [**1, 20, 21, 22, 25, 33, 34, 39**]. These essentially depend upon assuming that, for various institutional reasons, the government is not able to use traditional fiscal means to raise the level of savings in the economy to the optimal savings level, which is implicit in the maximization of the social welfare function subject to given resources and transformation constraints The privately generated level of savings is assumed to be socially sub-optimal, given that the private rate of time preference is higher than the social rate, as private savers are mortal and their altruism cannot be expected to extend to the infinite generations who are properly the concern of a society which (at least in principle) is immortal. Given a sub-optimal level of savings, and denied the use of traditional means for altering the savings-consumption balance, it is further assumed that, by influencing the choice of techniques and hence the resulting distribution of income accruals between profits and wages, the government can gradually alter the savings ratio towards the optimal one It is, however, not often noted that for this strategy to be successful, it must be assumed that the populace is subject to 'project illusion'. For otherwise the same factors which prevent the government from legislating the optimum savings level by more conventional means will also thwart any attempt at raising the savings level by this backdoor [1]

Furthermore, whilst the traditional formulations of the *SWR* have been much concerned with the differing social utilities of saving and consumption, and, implicitly, of effort and leisure, not much effort has been expended in taking account of the differing social utilities of income accruals to people in differing income groups within the same generation and/or as between different regions. Elsewhere I have presented a way of integrating the three relevant aspects of income distribution—inter-temporal, inter-regional and intra-regional—and deriving distribution weights for different income and/or regional groups (see [**17**]).[2] This essentially implies that there will now be a different s for each particular type of income recipient [3]

[1] For the theoretical problems in determining 's' see the references in p 114 n. 1 For empirical estimates of 's' for India, see my study of *Wells and Welfare* [**16**]

[2] Also see Stern [**35**]

[3] A referee has suggested that there are two other implicit assumptions in the traditional *SWR* model which are unrealistic, namely that the proportion saved out of profits is higher than out of wages, which may not in general be true, and that wage rates are independent of the techniques of production The first is not a necessary assumption for the traditional *SWR* model. For as the savings–consumption balance is sought to be affected by the choice of technique, the part of *profits* which are consumed should also be counted as a social cost

Normative judgements about the parameters of the social welfare function (for instance, the numerical value of the elasticity of social marginal utility) become indispensable.

What of the second policy parameter λ? Should the government value private disutilities of effort as social costs? The individualistic utility maximizing framework of economic theory conditions us to give an affirmative answer. However, I feel that there are a number of arguments which can be advanced for not putting a social valuation equal to the private valuation of the disutility of effort (up to a certain 'normal' range of hours worked per worker) in developing countries. First, these private disutilities are likely to reflect traditional attitudes to work, and other sociological features like relative caste status, traditional attitudes to particular occupations, etc., and one of the objectives of development policy is often to change these attitudes. Secondly, at low levels of living, governments may place a higher value on raising material standards of living than on increased leisure, and may want to raise the national product beyond the level implicit in private preferences.[1]

Thirdly, the higher standards of material consumption associated with higher incomes may have socially desirable qualitative effects in raising the over-all productivity of the labour force than an equivalent amount of 'leisure'. Fourthly, government actions and exhortations to induce people to donate their labour, either for no material rewards, or for rewards below the supply price of labour which reflects the private disutilities of effort (for example, the 'shramdan' movement in India) suggests that governments do, in fact, place no value on the private disutilities of effort within a certain range of 'normal' hours worked. Finally, by analogy with the postulated difference between the social and private valuations of savings in the economy, which accounts for 's' being greater than unity, there is no obvious *logical* reason why the private and social opportunity costs of the disutility of effort should be the same. This does not mean that govern-

as is the part of wages which is consumed. No new principles are therefore required to determine the social benefits of cases where a higher proportion of the profits of a project are saved, than wages. Neither is the second of the above assumptions necessary for the validity of the traditional *SWR* model. Even if the wage rate is specifically tied to the particular technique, the problem of evaluating the *SWR* still arises and the principles outlined in the text will be relevant. In fact the assumption of an institutional wage in the industrial sector tries to capture this linking of wages with particular 'techniques'.

[1] The implications for employment and aggregate output of putting a positive social valuation on these disutilities may be noted. If, for instance, the output forgone elsewhere in the economy with increased employment is zero, but labour can be hired only at a constant positive wage determined by a constant marginal disutility of effort, a social valuation equal to the private would mean providing employment and raising output up to the point where the marginal product of labour was equal to this wage, even though there would be no sacrifice in output elsewhere in the economy by expanding employment further, and total output could be increased till the marginal product of labour fell to zero. The latter would be the position which would be reached in this example if we did *not* put any social value on the private disutility of effort.

ments can *disregard* private preferences in either case. In the case of the private disutilities of effort, it cannot disregard individual preferences for income and leisure, for except in a totalitarian or slave economy, these will determine, first, any changes in output elsewhere in the economy when employment is increased (as we have noted in Section II) and, secondly, the actual minimum wage which will in fact have to be paid to get the labour for the newly created job. Assuming that all wages are consumed, the second factor will effect the macro-economic balance between consumption and investment in the economy, and if the social value of savings is greater than that of consumption ($s > 1$), there will in addition to the cost of the first factor (the output forgone) be the social cost of providing the increased consumption determined by the private supply price of labour. Thus, the private disutilities of effort will affect the value of the *SWR* indirectly, but in my view there is no reason for socially valuing these disutilities *per se*, and hence, in the various formulae derived in Section III, the social cost of the private disutility of effort λ should be put at nought.

University College, London

REFERENCES

1. BAUMOL, W. J., 'On the social rate of discount', *American Economic Review* vol. lviii, Sept. 1968.
2. BERRY, R. A., and SOLIGO, R., 'Rural urban migration, agricultural output and the supply price of labour in a labour surplus economy', *Oxford Economic Papers*, vol. 20, July 1968.
3. CHAKRAVARTY, S., 'The use of shadow prices in programme evaluation' in Rosenstein-Rodan, ed., *Capital Formation and Economic Development* (London, 1964).
4. DIXIT, A. K., 'Optimal development in the labour surplus economy', *Review of Economic Studies*, vol. xxxv, Jan. 1968.
5. ——— 'Short-run equilibrium and shadow prices in the dual economy', *Oxford Economic Papers*, Nov. 1971.
6. DOBB, M., *An Essay on Economic Growth and Planning* (London, 1960).
7. ECKSTEIN, O., 'Investment criteria for economic development and the theory of intertemporal welfare economics', *Quarterly Journal of Economics*, Feb. 1957.
8. FEI, J. C. H. and RANIS, G. *Development of the Labour Surplus Economy* (Homewood, Illinois, 1964).
9. GALENSON and LEIBENSTEIN, 'Investment criteria, productivity and economic development', *Quarterly Journal of Economics*, Aug. 1955.
10. HARBERGER, A. C., 'On measuring the social opportunity cost of labour', *International Labour Review*, 1971.
11. NURUL ISLAM, 'Concept and measurement of unemployment and underemployment in development economics', *International Labour Review*, Mar. 1965.
12. JOHNSON, G. E., 'The structure of rural-urban migration models' (mimeo), Institute of Development Studies, Nairobi, 1970.
13. JORGENSON, D. W., 'Surplus agricultural labour and the development of a dual economy', *Oxford Economic Papers*, Nov. 1967.

14. KAO, C. H. C., *et al.*, 'Disguised unemployment in agriculture: a survey', in Eicher and Witt, eds., *Agriculture in Economic Development* (New York, 1964).
15. KAHN, A. E., 'Investment criteria in development programs', *Quarterly Journal of Economics*, Feb 1951.
16. DEEPAK LAL, *Wells and Welfare* (OECD, Paris, 1972).
17. —— 'On estimating income distribution weights for project analysis', *IBRD Economic Staff Working Paper No. 130* (1972).
18. LEFEBER, L, 'Planning in a surplus labour economy', *American Economic Review*, vol. 58, June 1968.
19. LEWIS, W. A., 'Economic development with unlimited supplies of labour', *The Manchester School*, vol. 22, May 1954.
20. LIND, R. C., 'The social rate of discount and the optimal rate of investment: further comment', *Quarterly Journal of Economics*, May 1964
21. LITTLE, I. M. D., 'The real cost of labour, and the choice between consumption and investment', *Quarterly Journal of Economics*, Feb 1961
22. LITTLE and MIRRLEES, *Manual of Industrial Project Analysis, Vol. II, Social Cost Benefit Analysis* (OECD, Paris, 1969).
23. MAZUMDAR, D, 'The theory of urban underemployment in less developed countries' (mimeo).
24. MARGLIN, S., *Public Investment Criteria* (London, 1967).
25. —— 'The social rate of discount and the optimal rate of investment', *Quarterly Journal of Economics*, Feb 1963.
26. NURKSE, R., *Problems of Capital Formation in Underdeveloped Countries* (Oxford, 1955)
27. PAGLIN, M, 'Surplus agricultural labour and development. facts and theories', *American Economic Review*, vol. 55, Sept. 1965.
28. RAMSEY, F., 'A mathematical theory of savings', *Economic Journal*, Dec. 1928.
29. REYNOLDS, L. G., 'Economic development with surplus labour · some complications', *Oxford Economic Papers*, vol. 20, July 1968.
30. SCHULTZ, T W., *Transforming Traditional Agriculture* (New Haven, 1964)
31. SEN, A. K., *Choice of Techniques*, 3rd edn. (Oxford, 1968).
32. —— 'Peasants and dualism with and without surplus labour', *Journal of Political Economy*, vol 74, Oct 1966
33. —— 'Isolation, assurance and the social rate of discount', *Quarterly Journal of Economics*, Feb. 1967.
34. —— 'On optimising the rate of savings', *Economic Journal*, Sept. 1961.
35. STERN, N. H, 'Optimum development in a dual economy', *Review of Economic Studies*, Apr. 1972
36. STIGLITZ, J. E, 'Rural–urban migration, surplus labour and the relationship between urban and rural wages', *Eastern Africa Economic Review*, vol. 1, Dec 1969.
37. TODARO, M P, 'A model of labour migration and urban unemployment in less developed countries', *American Economic Review*, vol. 59, Mar. 1969.
38. HARRIS and TODARO, 'Migration, unemployment and development a two-sector analysis', *American Economic Review*, vol. 60, Mar 1970.
39. TULLOCK, G, 'The social rate of discount and the optimal rate of investment: comment', *Quarterly Journal of Economics*, May 1964
40. WONNACOTT, P, 'Disguised and overt unemployment in underdeveloped countries', *Quarterly Journal of Economics*, May 1962.

MARKETING CHARACTERISTICS AND PRICES OF EXPORTS OF ENGINEERING GOODS FROM INDIA[1]

By MARK FRANKENA

THE scholarly literature analysing the problems of exporting new manufactured goods from developing countries has concentrated on comparative costs of production, policies which discriminate between production for the domestic market and export, and trade barriers abroad Except in the product-life-cycle and two-gap models,[2] the role of export marketing in international trade has been ignored This paper presents data on export prices of Indian engineering goods which indicate that marketing may play an important role in determining comparative advantage and in explaining the difficulties encountered by semi-industrial countries in making a transition from import substitution to export in the case of new manufactured goods.

The importance of export marketing is suggested by the fact that the landed export prices of Indian engineering goods in 1967–70 were often substantially below those received for the same products and markets by competitors from advanced countries Contrary to the assumption typically made in discussions of new manufactured exports from developing countries, Indian engineering goods could not be exported at the 'world price' even though India was a marginal supplier

Export marketing problems

Marketing problems and practices appear to explain the fact that Indian engineering firms are not able to secure export orders at the landed prices received for the same products by competitors from advanced countries: India does not have a reputation for industrial production, established suppliers have advantages over new suppliers, and poor performance by many Indian exporters has created a bad reputation for Indian suppliers Apart from problems beyond India's control, even Indian engineering firms engaged in export have allocated few resources to export marketing

[1] I am grateful to Jagdish N Bhagwati for comments on a draft of this paper

[2] See R Vernon, 'International investment and international trade in the product cycle', *The Quarterly Journal of Economics*, May 1966, pp 190 207, S Hirsch, *Location of Industry and International Competitiveness*, Oxford University Press London, 1967, J R de la Torre, jun, 'Exports of manufactured goods from developing countries marketing factors and the role of foreign enterprise', *Journal of International Business Studies*, Spring 1971, pp 26–39, H. B. Chenery and M Bruno, 'Development alternatives in an open economy the case of Israel', *Economic Journal*, 1962, pp 73–103, and A MacEwan, *Development Alternatives in Pakistan*, Harvard University Press, Cambridge, Mass, 1971, pp 16–17, 57-61.

activities as a way of increasing export demand and prices, and the contribution of the Indian government in this area has been small. Like the East European countries, India has relied heavily on price concessions and to a significant extent on bilateral arrangements and tied financing to secure orders for manufactured goods facing marketing problems [1]

Export price discounts

The landed export prices of Indian engineering goods were compared with those received for the same products and markets by competitors from West European countries to determine the price discount, if any, required to sell Indian goods Comparisons were confined to prices at which exports actually occurred and do not include list prices set too high to compete Comparisons were restricted to hard currency exports without medium- or long-term credit, and they were made in third markets, not in the home country of the competing supplier

Table I presents the price discounts on Indian goods Our explanation for these discounts is the existence of the export marketing problems considered above However, while efforts were made to hold quality constant in comparisons, some of the discounts may reflect lower or more variable quality of Indian goods, particularly in regard to appearance Furthermore, in cases where comparisons were obtained from published sources rather than interviews by the author, it was not possible to confirm that the specifications and designs of the products compared were identical However, data from published sources were used only in cases where the source itself made an explicit price comparison based on interviews Because of the difficulty of assuring the accuracy of published reports, Table I lists the source of each comparison to distinguish those which were based on interviews by the present author

The data in Table I appear to justify two conclusions (i) Indian engineering goods other than commodity-like products were exported for hard currency only at prices below those received by competitors from West European countries, and (ii) the size of the price discount necessary to sell Indian goods was positively related to the marketing requirements of the product

Conclusion (i) is straightforward Not a single case was found in which Indian goods were exported at prices higher than those from Western Europe, and no case was found in which a non-commodity-like Indian product was exported without a price discount. The only products for which Indian

[1] Marketing problems and practices involved in export of Indian engineering goods are discussed in detail in M Frankena, 'Export of engineering goods from India', Ph D thesis, Massachusetts Institute of Technology, 1971, chapter VI

prices were reported to have been higher than those of any supplier other than a developing or East European country were stationary diesel engines, which were reported by published sources to have been priced 15 to 30 per cent above ones from Japan [1]

To test the hypothesis underlying conclusion (ii), the products were divided into three categories on the basis of their marketing characteristics: Group I, commodity-like products, which are standardized, bulk products for which marketing considerations like brand name and service play a negligible role, Group II, simple products, which do not require service but which are sold in smaller lots and are subject to brand considerations, and Group III, machinery for which brand and after-sales service are important [2]

While the difference between the means of the price discounts for Groups I and III is statistically significant at the 0·05 level, the differences for Groups I and II and Groups II and III are not However, most of the observations for which the discount is far from the group mean come from published sources (insulated wires, twist drills, tyres and tubes), which are less reliable *a priori* than interviews Using only data collected in interviews for the present study, all observations but one (dry batteries) conform to the following pattern: Group I, 0 to 10 per cent, Group II, 10 to 20 per cent, and Group III, 20 to 40 per cent.

However, this pattern of price discounts should be considered only an approximation in any case *A priori* there is no reason to expect discounts to have such a consistent pattern, since the discounts presumably vary with marketing input by the Indian exporter, degree of competition from suppliers from developing and East European countries, and the market share of the Indian exporter If demand for a product from India is less than perfectly elastic with respect to price even when India supplies a small share of the market, because there are marketing problems which make the Indian product an imperfect substitute for the product from advanced countries, one would expect the market share of the Indian product to be positively correlated with the price discount

Similar price discounts were found on exports from East European and other semi-industrial countries East European exports of consumer durables and machinery for hard currency have often been sold at landed prices 20 per cent or more below landed prices of goods from advanced

[1] National Council of Applied Economic Research, *Export Prospects of Diesel Engines*, New Delhi, 1967, pp. 27–8, and National Council of Applied Economic Research, *India's Export Potential in Selected Countries*, New Delhi, 1970, vol 1, pp 165–6

[2] For similar classifications of manufacturing industries based on the importance of advertising expenditures, service requirements, and captive distribution channels, see J S Bain, *Barriers to New Competition*, Harvard University Press, Cambridge, Mass, 1956, chapter 4, and J. R. de la Torre, jun, 'Export of manufactured goods from developing countries'.

TABLE I

Discounts below West European competitors' landed prices for Indian exports of engineering goods for hard currency

Product	Indian discount[1]	Competitor	Market
I *Commodity-like products*			
Steel bars and structurals	0 to negligible	Unspecified	Unspecified[2]
Steel tubes	7 to 10	Unspecified	Kenya[2]
Steel wire ropes	0 to negligible	Unspecified	Unspecified[2]
Power cables	0 to 2 5	W. Germany	Ghana[2]
Insulated wires	33	U K	Singapore[3]
II *Simple products*			
Hand tools	15 to 20	W Germany	Unspecified[2]
	12	Unspecified	Unspecified[4]
Twist drills	61	U K and W Germany	Denmark[5]
Dry batteries	3	U K	Ghana[2]
Light electricals, e g bulbs and their components	10 to 20	Netherlands	Unspecified[2]
Tyres and tubes	43 to 50	U K.	Iran[3]
III *Machinery products*			
Automobile parts	15 to 20	Unspecified (perhaps U.S)	Indonesia[6]
Sewing machines	25 to 40	France and Italy	Ghana and Nigeria[2]
Electric fans	5 to 36	U K	Australia, Kuwait, Iraq[3]
	50 to 62	U K	Iraq[7]
Bicycles	31	U K	Kenya[2]
	18 to 21	U K	Canada[8]
Stationary diesel engines	18 to 20	U K	Libya, Iraq[3]
	10	U K	Thailand[9]
Machine tools	20 to 30	Unspecified	U S , Canada[2]
	20 to 30	Unspecified	W Germany[10]
Unspecified machinery	20 to 25	Unspecified	Unspecified[3]

[1] Indian price discounts are expressed as a percentage of the competitor's price

[2] Information from interviews with Indian exporters, importers of Indian and West European goods in Ghana, Nigeria, Kenya, U S , and Canada, and Indian overseas trade representatives

[3] National Council of Applied Economic Research, *India's Export Potential in Selected Countries*, New Delhi, 1970, vol 1, pp 54, 148–51, 165–6, 182, and vol 2, p 30

[4] I Little, T Scitovsky, and M Scott, *Industry and Trade in Some Developing Countries*, Oxford University Press, London, 1970, p 194 The price discount of 12 per cent on Indian exports was derived from a statement by Little *et al* that the f o b. price of Indian exports was 17 per cent below the c i f price of Indian imports by subtracting a 5 per cent allowance for the difference between the f o b and c i f price of Indian exports.

[5] Engineering Export Promotion Council, *Market Survey Report on Selected Indian Engineering Products in Denmark*, Calcutta, 1968, pp 98–9

[6] *Eastern Economist*, New Delhi, 3 Apr 1970, p 655

[7] Indian Institute of Foreign Trade, *Electric Fans*, New Delhi, 1967, pp. 73–5.

[8] T K Sarangan, *Liner Shipping in India's Overseas Trade*, UNCTAD, Geneva, 1967, p 93

[9] National Council of Applied Economic Research, *Export Prospects of Diesel Engines*, New Delhi, 1967, pp 27–8

[10] Engineering Export Promotion Council, *West Germany's Market for Machine Tools*, Calcutta, 1968, p. 7.

Western countries.[1] Price discounts were also found for Latin American exports. It was reported that the Volkswagen subsidiary in Mexico planned 'to reduce the export price below that of German-produced models for sale in the American (US) southwest'.[2] Discounts of 20–30 per cent were reported for exports of Argentine bagging machinery and Brazilian paper-making equipment.[3]

Implications

The existence of export price discounts suggests that because of export marketing considerations there may be a substantial gap between the exchange rate at which the cost of production of non-commodity-like goods in developing countries would equal the c.i.f. price of imports and the exchange rate at which the same industries would be competitive in export markets, apart from the influence of transport costs and trade barriers.[4] This could help to explain the low level of new manufactured exports compared with domestic production, although policies which discriminate between import substitution and export are important.[5]

Moreover, since the level of the price discount appears to vary systematically with the marketing requirement of the product, export marketing requirements may influence comparative advantage.[6] Apart from the influence of costs of production, developing countries may have a comparative advantage in export of commodity-like products.

The existence of the price discounts discussed here suggests a problem for studies which attempt to measure the domestic resource cost of foreign exchange saved by import substitution. The calculations are sensitive to which international price is used.[7]

Three reservations concerning the preceding interpretation of the data on export price discounts on Indian engineering goods should be added. First, particularly in the case of machinery, one explanation for the observed discounts is probably the fact that Indian firms provided less back-up service to distributors than did their competitors from advanced

[1] Six examples are presented in M. Frankena, 'Export of engineering goods from India', Table VI.2.

[2] C. W. Reynolds, 'Changing trade patterns and trade policy in Mexico: some lessons for developing countries', *Food Research Institute Studies*, Stanford, 1970.

[3] *Business Latin America*, 17 July 1969, p. 230, and Business International Corporation, *The New Brazil*, New York, 1965, p. 30.

[4] Of course, to the extent there are economies of scale, exports could be made below average cost.

[5] It might also be necessary to reduce the domestic prices of Indian-made goods below the local market prices of imports from advanced countries, but India bans competitive imports.

[6] The relevant data would be for discounts as a percentage of value added.

[7] One might, for example, question Krueger's use of a mixture of c.i.f. prices of imports into India for some goods and f.o.b. prices of exports from India for others. See A. O. Krueger, 'The benefits and costs of import substitution in India: a microeconomic study', University of Minnesota, unpublished, dated October 1970.

countries. Consequently, the observed prices may not be for identical products, quite apart from matters such as reputation and brand name Since a lower level of service would reduce the private cost of exporting, and in all likelihood the foreign exchange cost as well, the price discounts in Table I may overestimate the role which marketing plays in determining comparative advantage and in explaining the difficulties encountered by India in exporting new manufactured goods. No attempt was made to determine the profitability of price concessions relative to marketing expenditures as an export strategy for Indian firms

Second, price concessions may be a temporary strategy for breaking into a new market and rapidly increasing the Indian share of the market. It could be hypothesized that price discounts will be reduced as foreign importers and users gain experience with Indian suppliers According to North American and Australian importers, that did occur for Japanese machine tools during the 1960s It is too early for evidence on this for India

Third, it has often been suggested that multinational corporations do not face the same marketing problems as domestic firms in exporting from developing countries If this is true, government policies toward multinational firms and the export policies of the latter would influence the comparative advantage in export and the export performance of developing countries [1]

University of Western Ontario

[1] See M Frankena, 'Restrictions on exports by foreign investors', *Journal of World Trade Law*, vol 6, no 5, 1972, pp 575-93.

ON THE THIRD WORLD'S NARROWING TRADE GAP: A COMMENT

By MANUEL R. AGOSIN[1]

In a recent paper in this journal,[2] L. Stein attempts to refute the hypothesis that there exists a structural external imbalance which constrains the growth of developing countries. He advances the following arguments in support of his view:

1. Contrary to the UNCTAD projections of an increasing trade deficit, the trade deficit of developing countries improved substantially between 1961 and 1966 and, therefore, came nowhere near the $12 billion mark associated with Dr. Prebisch's estimate of a $20 billion payments gap for the developing countries as a whole in 1970.[3]
2. The international reserves of developing countries have increased over the 1961–6 period.
3. The deterioration in the trade position of developing countries failed to materialize because of improved, and therefore poorly predicted, export performance, especially in the field of manufactures.
4. There is no necessary correlation between the rate of growth of imports and the rate of growth of GDP.

We would like to address ourselves to these points. But before entering into our main disagreements with Stein, a word about UNCTAD gap projections is in order. The original $20 billion payments gap (and the related $11 billion trade deficit—not $12 billion, as Stein believes) quoted by Dr. Prebisch was based on estimates prepared by the United Nations secretariat and contained in the 1963 World Economic Survey.[4] The intention of these estimates was to calculate the net capital inflows required by 1970 for the developing countries to achieve the target growth rate of 5 per cent during the sixties, a target which coincided with that of the United Nations First Development Decade. In addition, the projections were based on extrapolations of growth trends in the developed market economy

[1] The author is a staff member of the New York Office of the United Nations Conference on Trade and Development. The views expressed in this paper are his own and do not necessarily reflect those of the UNCTAD secretariat. The author wishes to thank Messrs Sidney Dell and V. K. Sastry for their extensive comments and suggestions.

[2] L. Stein, 'On the Third World's narrowing trade gap', *Oxford Economic Papers*, Mar. 1971.

[3] Raul Prebisch, *Towards a New Trade Policy for Development*, United Nations, New York, 1964, p. 5.

[4] 'Trade needs of developing countries for their accelerated economic growth', in *World Economic Survey, 1963*, United Nations, New York, 1964. For a description of the methods employed in obtaining the estimates, see *Studies in Long-Term Economic Projections for the World Economy*, United Nations, New York, 1964, Part II.

countries—and, consequently, in export growth rates of developing countries—during the 1950s.

Since these calculations were made, the UNCTAD secretariat has undertaken several projection exercises with a view to revising and extending to 1975 and 1980 the original gap estimates in the light of more recent information and an improved methodology [1] Whereas the original projections were based on a global model with one target rate of growth for all developing countries considered as a group, subsequent exercises have used individual country models and target GDP growth rates The use of individual models and targets which are established on the basis of each country's feasible range of economic growth rates tends to give greater realism to the projections finally arrived at. Another major improvement in the methodology relates to export projections While in the original gap exercise exports were projected at the global level by relating them to the rate of growth of the world economy, in subsequent exercises two independent export projections were obtained One set, built up from individual country exports, is basically demand-oriented and depends on the rates of growth assumed for the various areas of the world economy To serve as a check of the results obtained, a second set of export projections was built up from the commodity side taking into account the specific supply–demand factors for each major commodity. The two sets of export projections were then reconciled by making suitable adjustments [2]

A recent study by the UNCTAD secretariat compares the actual values for 1970 of imports, exports, savings, and investment with their projected levels obtained on the basis of the models used in the projections presented in TD/34/Rev 1 and the rates of growth of GDP observed during the 1960s [3] The difference between projected and actual magnitudes depends, of course, on the stability of the assumed parameters and the differences between the projected and actual values assumed for the exogenous variables The study shows that, at the regional level, the projected and actual values of the import surplus and the trade gap are reasonably close At the country level, the departures of actual from projected values, after making allowance for sampling error, are not more than what could be accounted for by departures in the exogenous variables from their assumed level Hence, the basic parameters used in the projections—e g the marginal propensities to import and save—show a reasonable degree of stability, and the methodology employed provides an adequate basis for projections

Because a projection, unlike a forecast, is a conditional statement

[1] See UNCTAD secretariat, *Trade Prospects and Capital Needs of Developing Countries* (TD/34/Rev 1, New York, 1968), and *Trade Prospects and Capital Needs of Developing Countries During the Second Development Decade* (TD/118/Supp 3, Santiago, 1972)

[2] For discussion of the methodology used, see TD/118/Supp 3, p 44

[3] TD/118/Supp 3

designed to indicate what the future value of an endogenous variable is likely to be if the assumptions regarding the evolution of other exogenous variables prevail, the fact that the *import surplus of developing countries in 1970 was less than the originally projected figure of $11 billion is not an indication that trade-gap calculations are futile exercises or, more important,* that developing countries are not facing a foreign exchange constraint to their economic growth. As Stein himself recognizes, the original gap estimates were conditional on the continuation of past trends in the developed market economies and the achievement of the target rate of growth of GDP of 5 per cent per annum in the developing countries The experience of the 1960s shows that the developed market economy countries as a whole achieved growth rates considerably in excess of the ones recorded during the fifties While the 1950-60 growth rate for such countries was 3·7 per cent, it increased to 4·9 per cent during the sixties [1] Partially as a consequence of this acceleration in the growth of their major trading partners the rate of growth of export volume of the developing countries jumped from 3·6 per cent during the fifties to 6·5 per cent during the sixties (6 0 per cent if major oil exporters are excluded) In addition, as Stein points out, the declining trend in the terms of trade of the developing countries was halted during the 1960s.[2] These factors alone largely account for a much lower trade gap in 1970 than originally projected.

Moreover, during the 1961–6 period, the developing countries were far from attaining the 5 per cent growth target. Stein compares the 4·2 per cent rate of growth achieved during this period to the 4·3 per cent rate recorded during the period 1950–2 to 1957–9 and concludes that the difference is too small to account for the significant departure from the projected trade deficit of $12 billion [3] It should be noted, however, that the relevant comparison is not with the growth rate experienced in the fifties The projected trade deficit was posited on a 5 per cent target rate of growth, and the actual rate of growth during 1961–6, as Stein is quite correct in indicating, came nowhere near that target, maybe partially as a result of a foreign trade structural constraint.

Even when all of these factors are considered, if major oil-exporting countries are excluded, the trade deficit of developing countries experienced an increasing trend during the 1960s. As can be seen in Table I, the trade deficit of all developing countries did decline substantially between 1960 and 1970, but the increase in the surpluses of major oil exporters more than accounts for the decline. If major oil exporters are excluded, the trade

[1] The calculation of these growth rates and other figures mentioned in the text are based on standard United Nations sources

[2] It should be noted that the declining-terms-of trade thesis is not strictly necessary for the existence of a structural external bottleneck, as Stein seems to imply. See Stein, p 117

[3] Stein, p 111

deficit of developing countries is shown to have increased substantially. As argued in one of the UNCTAD secretariat's publications,[1] *it is not legitimate to add surpluses and deficits. Apart from assuming the willingness of surplus countries to lend their surplus to deficit countries, this procedure is tantamount to equating surpluses on the trade balance with aid-giving capacity. Poor countries with a surplus, no matter how large, still have*

TABLE I

Trade deficit of developing countries

(in billion dollars)

	All developing countries	*Major oil exporters*[a]	*All developing countries, excluding major oil exporters*
1960			
Imports	29 0	3 1	25 9
Exports	26 1	6 1	20 0
Trade deficit	2 9	−3 0	5 9
1961			
Imports	29 7	3 1	26 6
Exports	26 3	6 4	19 9
Trade deficit	3 4	−3 3	6 7
1966			
Imports	40 1	5·0	35 1
Exports	37 9	9 5	28 4
Trade deficit	2 2	−4 5	6 7
1970			
Imports	55 0	6 8	48 2
Exports	54 0	13 9	40 1
Trade deficit	1 0	−7 1	8 1

SOURCE · IMF, *International Financial Statistics*, various issues.

[a] Bahrein, Iran, Iraq, Kuwait, Kuwait Neutral Zone, Libya, Muscat and Oman, Qatar, Saudi Arabia, Trucial States, Venezuela, and Yemen.

many unmet needs, and no valid argument could be constructed to suggest that they should lend their surpluses to other developing countries A share of the bias introduced by adding surpluses and deficits is removed when we exclude countries which are likely to have a structural surplus, as is the case of major oil exporters Because the original gap calculations made by the United Nations secretariat were based on a global model for all developing countries, it was not possible to distinguish between surplus and deficit countries. The use of country models in subsequent exercises has allowed

[1] TD/34/Rev. 1.

the UNCTAD secretariat to exclude surplus countries from the estimates of the developing countries' capital inflow requirements [1]

In order to discredit the notion of the foreign trade constraint, Stein presents evidence that the international reserves of developing countries have been increasing over the 1961–6 period [2] But this argument again proves to be fallacious Reserves are maintained to dampen the effects of

TABLE II

Reserves of developing countries

(in billion dollars)

	All developing countries	*Major oil exporters*[a]	*All developing countries, excluding major oil exporters*
1960			
Reserves	9 7	1 4	8 3
Imports	29 0	2 9	26 1
Reserves–imports ratio	0 33	0 49	0 32
1961			
Reserves	9 0	1 4	7 6
Imports	29 7	2 8	26 9
Reserves imports ratio	0 30	0 50	0 28
1966			
Reserves	12 1	2 6	9 5
Imports	40 1	4 4	35 7
Reserves–imports ratio	0 30	0 60	0 27
1970			
Reserves	18 2	4 1	14 1
Imports	55 0	6 0	49 0
Reserves–imports ratio	0 33	0 69	0 29

SOURCE IMF, *International Financial Statistics*, various issues

[a] Iran, Iraq, Kuwait, Libya, Saudi Arabia and Venezuela. No data are available for other major oil-exporting countries in the Middle East

short-run fluctuations in export earnings. Hence, an increase in reserves is not necessarily an indication of an improving long-run foreign exchange situation. As an indicator of the foreign exchange situation, the relevant magnitude is not the volume of reserves, but the ratio of reserves to imports. As can be seen in Table II, during the sixties this ratio remained constant for developing countries as a whole, and exhibited a downward trend if major oil-exporting countries are excluded

[1] This is done in TD/34/Rev 1, p 44, by excluding all surplus countries and in TD/118/Supp. 3, p. 70, by excluding major oil producers

[2] Stein, pp. 111–112

Stein makes a great deal of the fact that exports of manufactures from developing countries increased at a very fast rate during 1961–6.[1] However, the important question to ask is how significant are these exports in easing the external constraint of developing countries. Not only do they start from an extremely low base and will, therefore, take many years at high growth rates to make a dent into the foreign exchange problem, but in addition they are concentrated in a few countries which either enjoy a special situation *vis-à-vis* the developed countries (e g South Korea and Taiwan) or are among the industrially more advanced of the developing countries According to data compiled by the UNCTAD secretariat, in 1969 eight countries accounted for 67 per cent of all exports of manufactures from developing countries [2] These countries were Hong Kong, Taiwan, India, Yugoslavia, Mexico, South Korea, Brazil, and Argentina Hong Kong alone accounted for 23 per cent of the total

Stein's main argument is that there is no necessary correlation between the rates of growth of GDP and imports He ran a correlation between import and GDP growth rates for twenty countries over the 1961–6 period and found that the correlation coefficient was not significantly different from zero at the 5 per cent level [3] Since the correlation coefficient he reported was 0·77, this seemed quite odd to us We checked the computation and found that the coefficient is correct, but it is significantly different from zero even at the 1 per cent level

Apart from these statistical inaccuracies, there are more serious defects with his arguments. He shows that the correlation coefficient between GDP and import growth rates is of the same order of magnitude as the coefficient between GDP and export growth rates This is hardly surprising. the movement of imports normally tends to follow the movement of exports However, from the comparison of correlation coefficients and appealing to work done by Wall[4] and Emery,[5] Stein seems to draw the inference that the relationship between imports and GDP is not as important as the relationship between exports and GDP Apart from the fact that these conclusions cannot be demonstrated merely by showing that two correlation coefficients are roughly similar, the implications of this argument are not at all clear. If there is a strong correlation between exports and GDP growth rates, one must ask why such a relationship exists The explanation which emphasizes exports as a generator of demand is generally inadequate in the case of the developing countries, since their economic

[1] Stein, pp 114–15

[2] See UNCTAD secretariat, *Trade in Manufactures of Developing Countries, 1970 Review* (TD/B/C 2/102, Geneva, 1971), p 31.

[3] Stein, p 118

[4] D Wall, 'Import capacity, imports and economic growth', *Economica*, May 1968.

[5] R Emery, 'The relation of exports and economic growth', *Kyklos*, vol xx, 1967, fasc 2.

growth appears to be limited more on the supply than on the demand side. A more plausible explanation is that exports provide foreign exchange, and that the availability of imports is an essential factor constraining the expansion of the modern sector of the economy [1]

Of course, imports are only one factor constraining economic growth and hence the magnitude and significance of a simple correlation between GDP and import growth rates may not necessarily capture the strength of the association. If imports are an important factor in economic growth, we would expect that the partial association (as revealed by a multi-variable regression) between variations in GDP and import growth rates would be strong and positive With data for sixty-two developing countries,[2] the average ratio of investment to GDP and the rate of growth of imports were regressed on the rate of growth of GDP during the 1950–68 and 1960–8 periods The following results were obtained

1950–68

$$(1)\quad RY = \underset{(8{\cdot}14)}{0{\cdot}46}\,RM + \underset{(5{\cdot}81)}{2{\cdot}39} \qquad \bar{R}^2 = 0{\cdot}52$$

$$(2)\quad RY = \underset{(8{\cdot}78)}{0{\cdot}45}\,RM + \underset{(4{\cdot}03)}{0{\cdot}18}\,I/Y - \underset{(-0{\cdot}70)}{0{\cdot}58} \qquad \bar{R}^2 = 0{\cdot}62$$

1960–8

$$(3)\quad RY = \underset{(8{\cdot}29)}{0{\cdot}47}\,RM + \underset{(5{\cdot}97)}{2{\cdot}57} \qquad \bar{R}^2 = 0{\cdot}51$$

$$(4)\quad RY = \underset{(6{\cdot}93)}{0{\cdot}41}\,RM + \underset{(2{\cdot}45)}{0{\cdot}14}\,I/Y + \underset{(0{\cdot}36)}{0{\cdot}36} \qquad \bar{R}^2 = 0{\cdot}55$$

where

RY = average annual rate of growth of GDP,

RM = average annual rate of growth of imports,

I/Y = average ratio of gross fixed investment to GDP.

Figures in parentheses are t-ratios and $\bar{R}^2$ stands for the coefficient of multiple determination adjusted for degrees of freedom.

As can be seen from these equations, the rate of growth of GDP appears to be significantly related to the rate of growth of imports, the coefficient of RM is highly stable, and the coefficients of multiple determination are reasonably high considering we are dealing with a cross-section of many and diverse countries These results strongly suggest that export difficulties,

[1] As part of his argument, Stein (p. 119) contends that developing countries can always reduce luxury imports in order to increase the imports of machinery Anyone acquainted with developing countries knows that in most countries which have begun their industrialization process, imports of consumer goods in general have already been cut to the bone

[2] Obtained from United Nations, *Yearbook of National Accounts Statistics, 1969*, vol II The computer printouts of country data and regression results will be made available on request.

due either to slowly expanding world markets for primary products or to tariff and non-tariff barriers to the exports of manufactures, may indeed be a serious constraint on the economic growth of developing countries, via their effects on the imports of essential raw materials and capital goods.

UNCTAD, New York

ON THE THIRD WORLD'S NARROWING TRADE GAP: A REJOINDER

By L STEIN

WHILE M Agosin has rightly exposed some statistical inaccuracies in my paper, he has seriously misrepresented its central theme which was to examine 'factors influencing the less developed countries' (LDCs) trade gap over five years of the 1960 development decade'.[1] Accordingly, he has wrongly suggested that some of my arguments relating to this area were in fact mainly set up to refute the hypothesis that there exists a structural imbalance constraining growth

In my paper I showed that despite pessimistic forecasts, the LDCs were in fact able to decrease their trade deficit in the period 1961–6. Most of the discussion was concerned with analysing the factors responsible for this unexpected result [2] Firstly, I made mention of the fact the Prebisch and others believed that as long as the LDCs would grow, their trade deficits would deteriorate and that on the basis of an anticipated 5 per cent annual growth rate their 1970 deficit would be of the order of $12 billion [3] As growth did in fact take place in the period under review, one could not then argue that its absence explains the declining deficit I conceded that a growth rate of only 4·2 per cent occurred but none the less, as this barely differed from previous rates, one could hardly expect this factor to account for the *absolute* decline in the trade gap (In point of fact I had at no stage concluded, as Agosin claims I did, that the 4·2 per cent growth rate would still be appropriate for expecting the projected $12 billion deficit)[4] As the trade gap declined while growth proceeded, I then pointed out that it was unlikely that a reserve shortage unduly constrained imports since imports rose by 5·9 per cent per year In the same period reserves rose from

[1] L Stein, 'On the Third World's narrowing trade gap', *Oxford Economic Papers*, Mar 1971, p 110

[2] 'In 1969, for the second year in succession, the trade of the developing areas expanded very rapidly and brought the 1960–9 decade to a close with results that would have seemed unlikely, to say the least, at the decade's outset In 1960 the (f o b –f o b) trade account of the developing countries showed a deficit of $1 5 billion, projections made in the early 1960's indicated that this deficit might rise to more than $10 billion by 1970 Yet, by 1969, the exports and imports of the developing countries were almost in equilibrium', GATT, *International Trade*, 1969 (Geneva, 1970), p 135 The primary purpose of my article was to account for such an outcome

[3] Agosin corrects this figure to $11 billion

[4] Agosin refers the reader to p 111 of my article Part of the text states, 'This drop in the growth rate would obviously mitigate against the widening of the trade gap, but such a small degree of deceleration could hardly be expected to account for much of the trade deficit's decline' Only if one interprets the words 'trade deficit' to always mean the '$12 billion trade deficit' would Agosin's interpretation be correct I do, however, concede that since the '$12 billion trade deficit' was mentioned within the same paragraph, there is some scope for interpreting the passage in a different light to that intended

$9 billion to $12 billion with the reserve to import ratio remaining constant at 0·30. (The ratio was not explicitly stated but the absolute import and reserve figures were both contained in the article) As imports grew at anticipated levels, it seemed reasonable that one had to account for the trade gap decline in terms of an upsurge of LDC exports, and accordingly, five of the ten pages in the article were devoted exclusively to this end

One of my major points was that 'by far the most important factor in helping the LDCs' external balance was the rapid rise of economic activity in the Atlantic Community'.[1] (For some reason, Agosin gives the impression that this is something I have overlooked.)[2] Secondly, the unexpected rise of manufactured exports was deemed to be of crucial significance It is true that manufactured exports are, like oil, highly concentrated in a few countries but they currently account for approximately 24 per cent of total LDC exports and *if they had not grown at rates in excess of the world average during 1961–6, the LDCs would have experienced a growing trade gap* [3] Not surprisingly, I made 'a great deal of the fact that exports of manufactures from developing countries increased at a very fast rate'.[4] Finally, the failure of the LDCs' terms of trade to deteriorate was taken into account and here we have a good example of Agosin incorrectly summarizing my arguments. In a footnote on p. 135 he states that I seem to imply that the declining terms-of-trade thesis is necessary for the existence of a structural external bottleneck. The terms of trade discussion took place *entirely* in the context of an examination of factors affecting the trade gap (the central problem of the article) After noting Prebisch's reasons for believing in a worsening of the LDCs' terms of trade and offering some objections to them, this section ended with the following words 'Subsequent events have shown that in the period 1961 to 1966 the terms of trade have been fairly constant with the 1961–2 average nearly equalling that of 1965–6 Herein lies part of the answer as to why the LDCs' *trade gap* did not widen '[5]

To show that the LDCs' trade gap is not falling, Agosin insists on excluding oil exporters The logical basis for this is not clearly disclosed but I guess that it might be because these countries are not generally representative of the Third World. If this is indeed one of the reasons then one should perhaps also eliminate exporters of manufactures and a group of countries which have been able to achieve surpluses on account of possessing iron and

[1] Stein, p 114 [2] Agosin, p 135. [3] Stein, p 115

[4] Agosin, pp 138 and 139 Among his objections to my discussion of manufactures is the obvious fact that manufactures emanate primarily from 'the industrially more advanced of the developing countries'. In addition, Agosin on p 138 seems to regard the exports of manufactures of little significance 'in easing the external constraint of developing countries' yet in his concluding paragraph on p. 140 he suggests that along with primary products export difficulties of manufactures, 'may indeed be a serious constraint on the economic growth of developing countries'

[5] Stein, p 117 Italics not in the original

non-ferrous metals. However, this would result in a ridiculous situation in which the remaining countries would account for only approximately 44 per cent of all LDC exports If we are to look at the LDCs as a group, it seems unreasonable to exclude those that perform well to show that in reality the group is doing poorly and whereas I would agree that treating the LDCs as one homogeneous unit is far from satisfactory, it seems that the U.N and other bodies have previously adopted such an approach. Furthermore, it is not entirely irrelevant to consider surpluses attained by some LDCs while others are in deficit. Firstly, surpluses in some areas theoretically enable aid from rich countries to be rechannelled elsewhere and secondly, some large oil exporters have in fact been aiding other LDCs, e g Libya in relation to the U A R and Uganda, and Saudi Arabia in relation to some Arab states such as Jordan However, returning to the exclusion of oil exporters, Agosin's own table (Table I) shows that over the period which my article surveyed (1961–6) the LDCs, even excluding oil exporters, did not widen their trade gap

On the question of reserves, the LDCs' increase was not confined to 1961–6 but continued throughout the 1960s There is some general dispute regarding the validity of the use of reserve to import ratios as a relevant indicator of reserve adequacy.[1] Apart from the lack of any theoretical and satisfactory empirical framework establishing a systematic relationship between reserves and imports it is difficult to agree with Agosin on the basis of his own data in Table II that in the 1960s the ratio 'exhibited a downward trend if major oil-exporting countries are excluded'. The figures presented show that in 1960 the ratio for non-oil exporters was 0·32, in 1961 it was 0·28 and nine years later, in 1970, it was still 0·28 One could justifiably conclude that the over-all trend for the decade was fairly stationary Since it can be argued that reserves need not necessarily grow in the same proportion as transactions,[2] such a constancy of the reserve-import ratio might well indicate a reserve improvement

Turning to what has been described as my 'main argument , I unqualifyingly concede that in error I stated that no significant correlation existed between the growth rate of GDP and the growth rates of either imports or exports (I must admit that I was rather embarrassed in discovering that I had made such a mistake) Nevertheless, even when one corrects the error, the text that follows still stands, for as the strength of association between growth and imports and growth and exports could be of the same dimension, concentration only on the import-growth relationship may be too one-sided The export sector could be an economy's leading sector or it could be

[1] See, for example, F Machlup. 'The need for monetary reserves', *Banco Nazionale Del Lavoro Quarterly Review*, No 78, Sept 1966

[2] See, for example, F Machlup, *Plans for reform of the international monetary system*, Princeton University, 1964, p 17

stimulated by growth within the rest of the economy, or both, and contrary to Agosin's claim, it is not obvious that 'the explanation which emphasizes exports as a generator of demand is inapplicable to developing countries'[1] *for there are possible cases of LDCs giving vent to their surplus capacity* through exports.[2] While imports are needed for development, there are *possibilities that changes in capital to output ratios, the quality of the work* force, social relations, etc., could enable faster growth to occur at given import levels, and although 'imports of consumer goods in general have already been cut to the bone',[3] import substitution can still proceed over time.[4] For example, a country may import capital and some indispensable consumer goods which it is currently incapable of producing Ultimately some of the capital may enable it to produce some of the imported consumer items so that at a later period the composition of imports could change in favour of more capital goods

No one doubts that LDCs, like other countries, export in order to import, but what is at issue is whether the LDCs' exports can grow fast enough to assure them reasonable growth rates without necessarily incurring growing trade deficits The experience of the 1960s suggests that one could justifiably conclude that there is not necessarily a structural tendency towards external imbalance associated with the growth process As it happens, I share Agosin's view that the LDCs could doubtlessly grow faster if they were free to import goods without restraint but there is no basis for Prebisch's belief that LDC imports would necessarily have to rise at a faster rate than exports as growth proceeds.

Macquarie University, N S W.
Australia

[1] Agosin, p. 138

[2] See, for example, H Myint 'The "classical theory" of international trade and the underdeveloped countries', *Economic Journal*, June 1958

[3] Agosin, p 139

[4] Incidentally, I searched p. 119 of my article in vain for mention of 'luxury imports'

THE MARGINAL UTILITY OF INCOME

By COLIN CLARK

FOR a long time the search has been going on for some objective measurement of this concept, with little success

'Income' may be defined without valuation of leisure or other intangible advantages Income and leisure both have marginal utilities An hour's work has a marginal disutility, which contains two components, loss of leisure, and the effort of working. We may make the assumption that in the long run, after allowing people time to adapt themselves, the average duration of working hours of a community of given average income can be taken to indicate the point at which the marginal disutility of work to them equals the marginal utility of income.

We are concerned with the marginal utility of income, and not of 'money'. This latter concept calls for the valuation of a stock rather than of a flow per unit of time, and presents many further difficulties.

Professor Frisch opened up new territory in his paper in *Econometrica*, 1959, in which he showed that although consumption studies could not measure the marginal utility of money, they could measure the rate at which it declined with rising real income.

If x is real post-tax hourly income, and its marginal utility is y then Frisch's 'flexibility coefficient' is

$$\frac{dy/y}{dx/x}$$

He pointed out that both y and this coefficient must decline with increasing real income

In this paper, and in another in *Econometrica*, 1964, Professor Frisch (subsequently a Nobel Prizeman) indicates clearly how important it is to treat utility as a cardinal, measurable number, not merely as an ordinal, or ordering of preferences, as Sir John Hicks and Sir Roy Allen appear to have suggested. 'Sir Roy Harrod has declared that cardinal utility is necessary in dynamic analysis. With this I am in complete agreement. It is indeed the gospel which I have tried to preach for nearly a generation. The idea that cardinal utility should be avoided in economic theory is completely sterile . . . derived from one special and narrow part of theory viz. static equilibrium.'[1] Sir Dennis Robertson also always remained a strong 'cardinalist', and welcomed recruits to the Cardinal Club, for which indeed he offered to design a club tie.

[1] Frisch, *Econometrica*, 1964, p. 418.

A possible new source of information on the marginal utility of money arises from the studies by traffic engineers of 'valuation of time spent in travelling'. In preparing cases for road and bridge works, on which very large sums of public money are to be spent, the engineers endeavour to make an estimate of the economic value of the gains of the travelling public to be set off against the cost. While savings of the running costs of vehicles can be fairly precisely estimated, this is not the case with savings of travel time That it has an economic value is, however, not in doubt—travelling in a bus, or driving one's car during the rush hour, is not leisure, and any saving of time required for travelling constitutes a net addition to leisure. Originally some highly speculative estimates were made Then some ingenious methods began to be developed which depended on comparing travel charges on roads and bridges with the estimated savings, both in money and time, from using the toll route rather than the non-toll alternative This method depends on the assumption that the toll-fixing authorities have accurately gauged the optimum economic charge, which is probably only the case within wide limits [1]

Another method is to examine long-distance journeys for which alternative modes of travel are available, and to compare the time savings with the higher charges usually made for the faster mode of transport This method has the disadvantage that there are other attributes besides speed which may cause people to choose any particular mode of transport, and also that the type of traveller making the choice may be unrepresentative of the community as a whole.

An unpublished study by Miss F. M. Wilson[2] of comparative charges by air, first and second-class rail, ship, and long-distance bus for various journeys in the United Kingdom shows that, in comparing air journeys with second-class rail journeys, passengers pay about 0·6 £/hour saved. In comparison with first-class rail journeys, however, the result comes out at 0 4 £/hour saved. The 0 2 £ saved difference may be assumed to be a measure of the greater comfort of first-class travel Air travel, on the whole, is as uncomfortable as second-class rail travel, and the larger figure may be taken as representative

The passengers who make this choice will, however, on the whole be fairly high-income business men

In the choice between air and sea travel to Belfast or to the Orkneys and Shetlands, the result works out at 0 4 £/hour. This will not be such a high-income group. On the other hand, in travel to the Channel Islands

[1] The present writer conducted one of these early surveys, in fixing the toll for the Story Bridge in Brisbane in 1939 It soon became apparent that the toll had been fixed at an uneconomically low level

[2] University of Sussex, Institute of Development Studies Values are those of 1966

and the Isle of Man, the difference is valued at only 0 1 £/hour This latter, however, is largely holiday traffic, and for many travellers the sea voyage is part of the enjoyment.

The choice between second-class rail and long-distance bus is one made, on the whole, by low-income travellers. The relationship appears non-linear Savings up to $2\frac{1}{2}$ hours are valued at about 0·15 £/hour, above that at 0 3 £/hour. This non-linearity may be explained by increasing fatigue,

Fig. 1

or more plausibly by the fact that the bus gains some comparative advantage from its greater number of stopping-places but that on longer journeys this advantage becomes relatively less important

While there are clearly many difficulties about such a method, it gives us some idea of orders of magnitude, and suggests also that higher-income travellers are willing to pay more per hour saved.

In Australia comparison of air, bus, and rail fares for long journeys undertaken (without the use of sleeping-cars) indicate a 'trade-off' of about \$1 per hour. The relationship appears to remain approximately linear, even for very long distances. Train travellers in Australia now, however, include a high proportion of pensioners and others of low income. Higher-income travellers would probably place a higher valuation on time.

This method of analysis was placed on a more secure foundation in a highly original study by Beesley.[1] Arising out of proposals to move a large number of civil servants in the Ministry of Transport, of all grades, from one workplace to another, in London, and the preferences expressed by them it was possible to obtain a 'trade-off' between money and saving in travel time. For three income groups the values ranged from 0·1 £/hour for the lowest to 0·55 £/hour for the highest, the valuation of time (i.e. of net addition to leisure) being much higher, relative to income, in the higher income groups.

An original and thorough study by Thomas[2] indicated that all previous studies had much underestimated the value of travel time savings and gave for the United States an average as high as 2 8 dollars/hour Thomas subjected the results of a number of traffic studies to an elaborate procedure whereby $p(x)$, the probability of a car taking a non-toll route, was expressed as a logit function of $f(x)$, an expression of 'motorist and route characteristics' In the most successful procedure, the 'characteristics' were confined to the toll, the time loss on the alternative route, and the family income, thereby making it possible to estimate the 'trade-off' between time and money at any given income level. (In another investigation attempts were made also to analyse the effects of the age of the car, and the sex of the driver!)

The equation used was

$$p(x) = e^{f(x)}/(1+e^{f(x)}),$$

and iterations were undertaken until the best fit was obtained

For Sweden, a comparatively high-income country, an estimate of 1 5 dollars/hour has been made [3] In Sweden likewise there had been a tendency to underestimate, and many traffic experts were reluctant to accept this figure

These results are in the form of the amount of income traded against an hour's saving of time, or addition to leisure We may invert the expression, treating the discomfort or effort of an hour's travel by bus or by car (the two being assumed to be much the same), in comparison with genuine leisure, as an objective unit of disutility The reciprocal of these expressions therefore gives us the marginal utility (for the traveller is taking his 'trade-off' at the margin of his income) of one unit of income, in terms of 'travel-hour' units of utility

For comparison, all incomes have to be expressed in the same unit of real income, which we take as dollars of 1958 purchasing power/earner/year.

[1] *Economica*, May 1965

[2] *The Value of Time for Passenger Cars*, Stanford Research Institute, Project 5074, 1967, vol 2

[3] Classon, 1963, quoted *Traffic Quarterly*, July 1965, p 429.

This unit was chosen because 1958 is the base for real income calculations in the U.S. national accounts.[1]

Goodwin[2] has assembled considerable further information published in the U.K., and has generously permitted its use in this paper. Goodwin also takes the reciprocal of the 'value of time', i.e. the marginal utility of money in terms of 'travel-hours' or their equivalents, as his object of measurement. He seeks to relate this to the reciprocal of income, i.e. implicitly assuming that the reciprocal measures the marginal utility of income. While recognizing that a relationship somewhat akin to this may prevail, further investigation is desirable. Goodwin also makes the important point that an hour's leisure is an objective piece of welfare, as valuable to a poor man as to a rich man—though the latter will offer more money in order to obtain it. Many traffic studies have mistakenly treated an hour's time saved for a car-traveller as worth more than an hour saved for a bus-traveller—and traffic plans have been distorted in consequence in favour of car-travellers.

When information on hourly earnings is not available, it is assumed that annual incomes quoted represent 2,000-hour working years.

It is, however, also necessary to make adjustments for taxation. Everyone should, and indeed does, plan his expenditure and savings in relation to his post-tax, not his pre-tax income. This adjustment can be made only approximately.

It is interesting to observe that international comparisons show similar relations to intra-national comparisons. It might be contended that in a richer country, the desire to emulate other men's consumption might raise the marginal utility of income, i.e. lower the valuation placed on leisure at any given level of real income, in comparison with the same level of real income in a poorer country. But it appears that this does not have a significant effect.

We may now return to Frisch's 'flexibility coefficient', defining x (real post-tax hourly income) in U.S. \$ of 1958 purchasing power, and y (marginal utility of income) in 'travel-hour' units.

For contemporary Norway Frisch thought that the coefficient $\frac{dy/y}{dx/x}$ should be about -2 on the average. However, he speculated that it might range from -10 'for an extremely poor and apathetic part of the population' to

[1] Thomas's figures refer to 1965–6, when the general price level was 1.123 on 1958 base. The figure of 1.5 dollars quoted for Sweden was presumed to represent 7.8 kroner (at current exchange rate), which was linked to U.S. purchasing power on comparisons published in *Wirtschaft und Statistik*, Mar. 1955. British prices were linked by means of the Gilbert-Kravis study, *An International Comparison of National Products* (O.E.C.D., 1957), for 1950. Between 1950 and 1963, the year to which Beesley's study referred, average prices of consumer goods and services (from *National Income and Expenditure*) had risen by a factor 1.539. This index is also used to adjust U.K. data for other years.

[2] *The Value of Time and the Distribution of Income* (unpublished), Research Group in Traffic Studies, University College, London, Nov. 1970.

TABLE 1

The Beesley, Thomas, and Swedish results with the additional information quoted by Goodwin are tabulated below, and the results plotted on Fig. 2

Authors	Periodical reference	Date of observations	Hourly wage in new pence: As given	Hourly wage in new pence: Post tax	In $ of 1958 purchasing power[a]	Value of one hour's time in new pence as given	Do in $ of 1958 purchasing power (reciprocal)
Lee and Dalvi	Manchester School Sept 1969	1966	53.35	49.0	1.625	16	1.87
Rogers Townsend, and Metcalf	Local Government Operational Research Unit Report 67		68.5	59.0	2.08	16	1.77
Barnett and Saalmans	Greater London Council P/209 Jan 1967	1964	146	102	3.66	20	1.39
			65–101	57–79	2.05–2.82	15	1.84
			35.5–46	33.5–42.5	1.21–1.53	10	2.77
Quarmby	*Journal of Transport Economics and Policy*, Sept 1967	1964–6	115	88	3.04	24	1.20
			75.5	63.3	2.19	26	1.11
			52.5	47.4	1.64	13	2.22
			36	34	1.18	8.75	3.30
Wabe[b]	*Journal of Transport Economics and Policy*, Sept 1969	1951	20	19	0.99	10	1.92
		1961	35	33.2	1.31	16	1.58
Mansfield[c]	*Journal of Transport Economics and Policy*, May 1969	1966	43	39.9	1.32	16	1.88
Stopher	London University Ph D thesis, July 1967	1967	149	105	3.40	62.5	0.49
			130	96.5	3.12	27.5	1.12
			87.5	171	2.30	20	1.55
			62.5	54.9	1.78	16.25	1.84
			35	33.2	1.08	11.25	2.72

[a] Post-tax wage first adjusted to 1963 level by National Accounts deflator for consumption then converted to U S $ of 1958 purchasing power by factors based on the Gilbert–Kravis comparison for 1950 (O E C D)

[b] Calculations extended by Goodwin—males only

[c] Recreation travel to Lake District

Authors	Periodical reference	Date of observations	Hourly wage in new pence: As given	Hourly wage in new pence: Post tax	In $ of 1958 purchasing power[a]	Value of one hour's time in new pence as given	Do in $ of 1958 purchasing power (reciprocal)
Other data							
Beesley	*Economica*, Mar 1965	1963	120	91	3.38	56	0.48
			42.5	39.4	1.47	16	1.69
			32.5	31.2	1.16	10	2.70
Thomas	See text	1965–6	12,000[d]	10,525[d]	4.74		0.35
			10,000[d]	8,886[d]	3.98		0.40
			8,000[d]	7,228[d]	3.21		0.96
Sweden		1963			3.5[e]		0.68

[d] Family income $ year See p 90 of Thomas's report for analysis of income effect

[e] Highly approximate estimate for supposed car-own[illegible] income distribution [illegible]

as low as —0·1 'for the rich part of the population with ambitions toward conspicuous consumption'. So far we have considered only an income range of 1–4 $/hour (at 1958 purchasing power), and the shape of the relationship might change at extreme values. However, Frisch's range of expected values does appear improbable. Pearce[1] thinks that British data indicate a figure of —2. Barten[2] obtained a figure of —2·1 for the Netherlands, but later[3] raised his estimate to —3·1. Powell, Van Hoa and Wilson[4] found —1·5 for U.S.A. For Australia, Powell[5] found —2·35 and Van Hoa[6] —2·46.

Frisch's coefficient (with sign reversed) is clearly a declining function of real income. A relationship is fitted on the estimated values of this coefficient for U.S.A., Australia, and the Netherlands, whose disposable personal incomes per working hour after tax in $ of 1958 are estimated at 2·58, 1·63, and 0·81, respectively. The British and Norwegian figures appear low, for these countries' income levels. These were only rough estimates based on less careful analysis.

Frisch's coefficient (with sign reversed) is clearly a declining function of real income. It is tabulated below in relation to real incomes per hour. Incomes were defined as personal disposable income (i.e. exclusive of direct taxes payable, inclusive of social welfare benefits)—sometimes this had to be estimated from related data. These were converted into real terms for each country by the consumption price deflator and converted to U.S. dollars on data for consumption price comparisons (not whole national product deflators), on the European (Australian) weights, in Gilbert and Kravis, *An International Comparison of National Products and the Purchasing Power of Currencies* (O.E.C.D., Paris, 1957 edition), and Haig, *Real Product Income and Relative Prices in Australia and the United Kingdom* (Australian National University Press, 1968).

This was divided by the whole employed labour force, and the average annual hours worked.

We then have

	Period covered	r	$-\frac{dy/y}{dr/r}$
U.S.	1949–63	2·58	1·5
Australia	1949–50 to 1961–2	1·63	2·35
U.K.	1955–9[a]	1·11	2
Norway	1950	0·86	2
Netherlands	1922–39, 1950–61	0·81	3·12

[a] Pearce refers only to 'the past five years'.

[1] *Econometrica*, 1961.
[2] Ibid., 1964.
[3] Ibid., Apr. 1968.
[4] *Southern Economic Journal*, Oct. 1968.
[5] *Econometrica*, July 1966.
[6] *Australian Economic Papers*, Dec. 1968.

No function can at present be fitted satisfactorily to this relationship. A linear fit is impermissible because it would eventually give, at high incomes, a point of inflexion beyond which marginal utility could increase with increasing income. An exponential (semi-logarithmic fitting) does not prove as good as a power (logarithmic) fitting

$$-\frac{dy/y}{dx/x} = 2{\cdot}35x^{-0\cdot412}.$$

If
$$-\frac{dy/y}{dx/x} = ax^{-n},$$
then
$$dy/y = -ax^{-n+1}\,dx,$$
and
$$\log_e y = \frac{a}{n}x^{-n}+C$$
$$= 5{\cdot}71x^{-0\cdot412}+C$$

The data so far obtained were grouped for fitting a relationship.

Number of data grouped	4	4	3	4	5	4
Average x	1·10	1 37	1 68	2 25	3 23	3·97
Average y	2 66	1 98	1 94	1 57	0 85	0 70

In order to fit the function at the lowest end of the scale, the arbitrary assumptions were made that the minimum value of x was 0·333 (indications of money incomes are lower but we must take into account the large amount of income received in kind in peasant communities); and also that at this level of real income people are willing to trade as much as 15 hours' leisure against one hour's earnings (i.e 45 hours against 1 $ of 1958 purchasing power

The constant is fitted as follows

x	0 333	1 10	1 37	1 68	2 25	3 23	3 97
$5{\cdot}71x^{-0\cdot412}$	8 99	5 49	5 01	4 60	4 09	3 53	3 25
$\log_e y$	3 80	0 98	0 68	0 66	0 45	−0·16	−0 36
C	−5 19	−4 51	−4 33	−3 94	−3 64	−3 69	−3 61

The arbitrary initial value is omitted and the constant averaged at −3·95. The expression $\log_e y = 5{\cdot}71x^{-0\cdot412}-3{\cdot}95$ is plotted on Fig 2, together with the grouped data It must be regarded as highly provisional.

Another possible source of international comparison of the valuations placed upon leisure might be found in payments to film 'extras', who have to wait about all day, and may have to don strange costumes, but do no real work—they are just being paid for loss of leisure. In California[1] the 1969 rate was 29 $/day, which is almost equal to the average wage. A film producer willing to work in the poorer state of West Virginia could, how-

[1] *Wall Street Journal*, 27 June 1969.

ever, obtain extras at 17 $/day, a lower proportion of the current wage than in California. This field should be further examined.

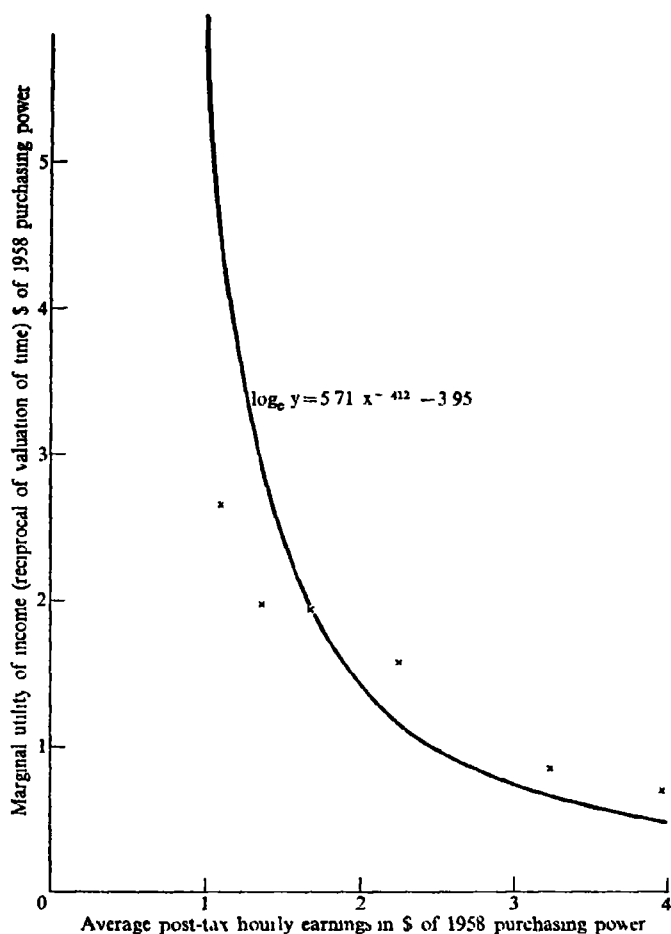



FIG 2

We now consider information on average hours worked at various real income levels, with the object of using this equation for marginal utility of income To analyse the disutility of labour the data used are national averages of weekly hours worked in relation to average real income/person/year.[1] The latter is converted into income/earner/year by dividing by a factor of 0 37. To convert weekly to annual hours we require information

[1] Current information on hours from International Labour Office Year Book of Labour Statistics, and on income from U N National Accounts Statistics Earlier information from *Conditions of Economic Progress.* See Appendix for some further data

TABLE II

Basic data on hours and incomes

		Average weekly hours	*Income per head $ of 1958*	*Weeks/year*	*Hours/year*
Australia[b]	*1891*	48.8	832	51.1	2,494
	1913–14	47.3	850	51.1	2,417
	1928–9	43.1	1,011	50.3	2,168
	1958–9	39.9	1,877	49.1	1,959
Austria[d]	1960	45.5	1,570	51.1	2,325
Belgium[a]	1960	46.4	1,494	51.1	2,371
Canada[d]	1960	40.4	1,656	49.4	1,996
Columbia[d]	1960	51	393	51.1	2,606
Denmark[a]	1960	44.8	1,709	51.0	2,285
France[e]	1848	72	251	51.4	3,701
	1885	66	359	51.4	3,392
	1896	61.2	445	51.4	3,146
	1929	50.0	791	51.1	2,555
	1960[a]	47.5	1,643	51.1	2,427
Germany[e]	1885	66	423	51.4	3,392
	1913	57.8	636	51.4	2,971
	1960[a]	46.8	1,823	51.1	2,391
Italy[e]	1885	76	207	51.4	3,906
	1907	60	255	51.4	3,084
	1960[a]	45.6	1,135	51.4	2,330
Japan[e]	1901	72.5	206	51.4	3,726
	1926	59	430	51.4	3,033
	1960[c]	46.8	837	51.1	2,391
Netherlands[a]	1960	48.2	1,470	51.1	2,463
New Zealand[e]	1938–9	41.0	1,500	49.5	2,030
Norway[a]	1960	45.4	1,289	51.1	2,320
Peru[c]	1960	46.0	322	51.1	2,351
Spain[c]	1960	43.6	830	50.5	2,202
Sweden[f]	1861–5	63.5	162	51.4	3,264
	1891–5	62.1	288	51.4	3,192
	1911–15	57.3	590	51.4	2,945
	1936–9	48.1	1,113	51.1	2,458
	1960[d]	38.4	2,142	48.5	1,862
Switzerland	1960[d]	46.1	2,084	51.1	2,356

SOURCES

[a] Dennison, *Why Growth Rates Differ* (Brookings Institution), p 55

[b] Commonwealth Year Book and Labour Report

[c] International Labour Year Book

[d] Do — Manufacture

[e] Conditions of Economic Progress

[f] Bagge, *Wages in Sweden 1936*, p 48

[g] Hours of weighted composite as given by Barger, *Distribution's Place in American Economy*, except that in place of the uniform 51 hours there given for agriculture extrapolated hours from Barger and Lansberg *American Agriculture* are used

TABLE II (*cont*)

		Average weekly hours	*Income per head $ of 1958*	*Weeks/year*	*Hours/year*
U K[e]	1836	66	303	51·4	3,392
	1886	55	822	51·1	2,810
	1913	53·5	1,089	51·1	2,734
	1929	45·3	1,115	51·1	2,315
	1960[a]	44·9	1,314	50·1	2,249
U S A[e g]	1850	62·0	400	51·4	3,187
	1860	60·6	325	51·4	3,115
	1869	57·8	325	51·4	2,971
	1879	57·0	600	51·4	2,930
	1889	55·8	729	51·4	2,868
	1899	54·6	845	51·1	2,790
	1909	52·9	1,042	51·1	2,703
	1919	50·0	1,165	51·1	2,555
	1929	49·0	1,507	51·1	2,504
	1940	43·8	1,546	50·4	2,208
	1960[k]	41·3	2,430	49·2	2,032

on public holidays, paid vacations, etc., which is scarce and conflicting.[1] The procedure adopted is to multiply the figure of hours/week when it exceeds 55 by a factor of 51·4 (i.e. assuming only 4 days/year holiday); hours/week of 45–55 by a factor of 51·1, and for hours/week below 45 to reduce the number of weeks/year by 0·4 for each hour/week below 45. Thus a 40-hour week is assumed to be associated with a 49·1-week year.

The available information on incomes can be approximately grouped to show the median levels of real income corresponding to certain hours. (It is assumed that the economies studied are in long-run equilibrium in this respect, and that people as a whole are not working either longer or shorter hours than they are willing to work, given their real incomes.) The marginal utility of income at each of these levels is calculated by the above formula, and assumed to represent the marginal disutility of effort plus loss of leisure. Personal post-tax income differs from national income by the amount of personal taxation, undistributed company profits and income of government enterprises *less* transfer incomes redistributed. Pending further inquiry, a downward adjustment of 15 per cent is made in the two highest income categories and 10 per cent in the next to allow for these factors. They are of little importance for the lower-income countries or for data further in the past.

We estimate a minimum subsistence income of 300 $ of 1958 purchasing power/earner/year. On the face of it, the national accounts of some countries (and some early data) indicate considerably lower average

[1] Dennison in *Why Growth Rates Differ* has attempted to assemble the available contemporary information for United States and Western Europe.

incomes. Only a highly approximate comparison of purchasing power with the higher-income countries is possible; but we must remember (1) that in the higher-income countries food carries the cost of transport and distribution, whereas in the low-income countries most of it is valued at the farm, and a substantial imputed addition should be made for transport and distribution costs to render the value of food consumption comparable with the higher income countries, (2) in the lower-income countries the purchasing power of money over many commodities, particularly services

TABLE IIA

Data of Table II grouped by income

Median incomes $ of 1958/ year	*Hours/ year*	*$/hour*	*Do adjusted for taxation*	*Marginal utility of $1*	*Marginal utility of hourly earnings*	*Marginal disutility of effort*
6080	2,095	2 90	2 47	0 99	2 44	1 44
4860	2,232	2 18	1 85	1 65	3 05	2 05
4450	2,300	1 94	1 74	1 84	3 20	2 20
4050	2,380	1 70		1 90	3 23	2 23
3240	2,455	1 32		3 16	4 17	3 17
2835	2,555	1 11		4 47	4 96	3 96
2225	2,505	0 89		7 76	6 90	5 90
1845	2,880	0 675		16 0	10 77	9 77
1620	2,928	0 553		28 5	15 8	14 8
1150	3,130	0 367		109 6	40 1	39 1
	3,440	0 333		151 4	50 5	49 5

and buildings, is very much higher than is indicated by its exchange rate (3) Many goods and services are supplied by the non-monetized sectors (apart from agriculture, whose product has been fully taken into account). In comparing Thailand—admittedly an extreme case, being a country where internal prices are kept low by the export tax on rice—with U K Usher[1] found the level of real income as much as 4½ times that indicated by an exchange rate comparison

We assume that on the average working hours are adjusted to the point where their marginal disutility is matched by the marginal utility of income. For each hour worked there is the disutility of one hour's loss of leisure, with a further disutility representing the effort required This latter rises approximately linearly with hours worked

The present equation for marginal utility of income will no longer serve for some of the higher incomes of the future. Not only the marginal utility of $1, but the marginal utility of one hour's earnings, may be expected to continue to fall with rising real income—if not, we should expect to find men working longer hours as their incomes rose. It is, however, probable that the rate of fall of marginal utility of one hour's earnings—and hence in the number of hours worked—will become very

[1] *Economica*, May 1963.

small. A Japanese estimate[1] shows average per head income in the year 2001 at 2·5 million yen of 1960 purchasing power, equivalent to per worker earnings (post tax) of about 17 dollars of 1958 purchasing power per hour associated with a fall in working hours only to 1,600 per year At this high level of income the marginal utility of $1 is calculated from the equation at 0·114, and of an hour's earnings therefore at 1·94 The curves for marginal disutility of effort, however inaccurate they may be for the low-income-long-hours positions, show clear signs of flattening to a very

TABLE III
Swedish Data

	NNP $ of 1958 billions	*Employment millions*	*Income/worker of $1958/year*	*Hours*	*Hourly income*	*Marginal utility from equation*		*Marginal disutility of effort*
						$1	*1 hour's earnings*	
1901–5	1·844	1·747	1,057	3,051	0·346	133·4	46·2	45·2
1906–10	2·225	1·805	1,232	2,968	0·415	70·0	29·0	28·0
1911–15	2·628	2·012	1,308	2,990	0·437	58·9	25·7	24·7
1916–20	2·945	2·239	1,315	2,924	0·449	54·7	25·7	24·7
1921–5	3·065	2·280	1,343	2,500	0·537	30·5	16·4	15·4
1926–30	3·77	2·472	1,527	2,580	0·593	23·0	13·6	12·6
1931–5	3·87	2·530	1,528	2,460	0·621	20·0	12·4	11·4
1936–40	4·95	2·76	1,791	2,425	0·740	12·5	9·25	8·25
1941–5	5·56	2·87	1,937	2,480	0·781	10·7	8·35	7·35
1946–50	7·73	2·96	2,612	2,190	1·191	3·89	4·64	3·64
1951–5	9·53	3·00	3,177	2,128	1·400	2·46	2·67	1·67
1956–60	11·32	3·05	3,710	2,032	1·825	1·66	3·03	2·03
1961–5	14·08	3·295	4,265	1,900	2·245	1·12	2·52	1·52

Real gross product linked on purchasing power of krone 0·274 $ in 1950

Net taken as 10% less Since 1950 U N Year books—previously Johannson spliced on 1950 5

Hours to 1913 compared from ratio of hourly to annual earnings (Definition I) p 48 of *Wages in Sweden*, subsequently I L O Year Book (Uniform multiplier of 50 replaced by new variable) Employment Conditions of Economic Progress to 1945, then O E C D

slight slope as annual hours fall below 2,000. That in a wealthy future community marginal disutility of effort, when 1,600 hours/year are being worked, might stand at about 0·94 units, as indicated by the above relationship, seems quite probable

A continuous series (which has been included in the above general averages) is available for Sweden. But at the lower income levels the disutilities appear to be greater (hours less) than would have been expected from the equation

It may be objected that many peasant farmers in Asia, at very low income levels, nevertheless have a short working year. This, however, is unavoidable in view of the climate, and the smallness of their land and capital resources. I am content to accept the assurance of the Indonesian economist Adiratma Roekasah[2] that they put little or no value on this

[1] Suzuki, 4th International Symposium on Regional Development, Japan Centre for Area Development Research 1972, p 9

[2] Private communication.

enforced leisure, and that they would much rather be working, even for a very low remuneration. Indeed, when offered wage work they are willing to work for very long hours.

It may be a different matter for African tribesmen, whose primitive 'cut and burn' agriculture may occupy them for only about 1,000 hours

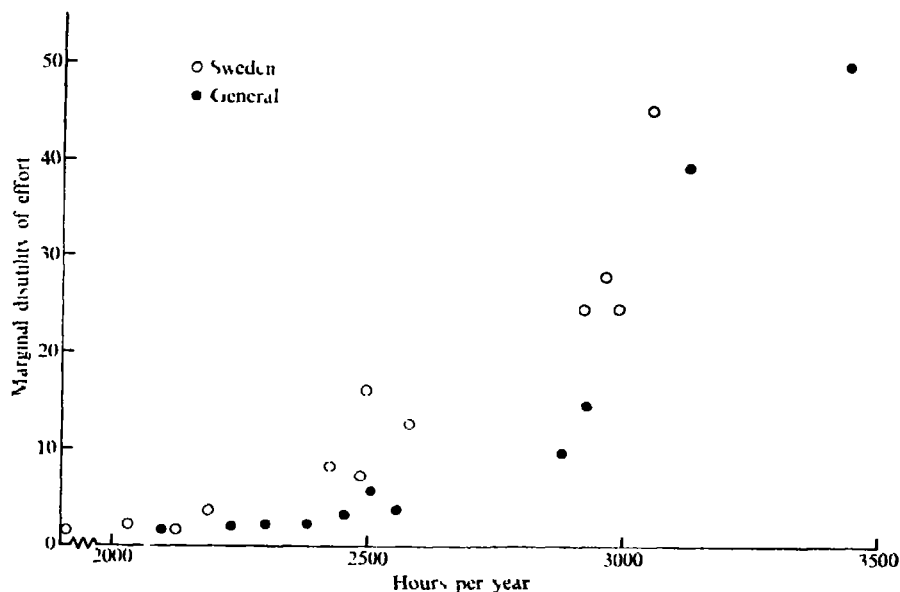



FIG 3

per year, though they work longer if house repairing, public duties, and other activities are taken into account But they are largely cut off from the commercial world (principally by lack of transport) and cannot therefore be said to have an effective marginal utility of money income

APPENDIX

Henle (Monthly Labor Review, 1967) gives the following data for U.S A

Average weekly hours

	1900	*1930*	*1948*	*1956*	*1960*
Agriculture	67	55			
Other work	56	43			
All workers			43 4	41 6	40 8
Do. excluding part-time workers			46 8	46 0	45 5

Average vacations and other paid holidays were

	1910	1948	1952	1956	1960
Average vacation weeks/year	0 3	1 0	1 1	1 2	1 5
Other holdays days/year	1				7
Other holidays weeks/year	0 36				1 1

For comparison the following estimates can be made for Britain

	1831–71	1871–1938	1960
Average hours/week	60	47	45
Average vacation and public holidays weeks/year	0 67	1 0	2 7

Kreps[1] shows weekly hours of work in U S A falling from 61 9 in 1890 to 40 5 in 1964, with 4 days/year public and 6 days/year other paid holidays in the latter year.

After French hours/day had been limited to 10 in 1900, and Sunday rest introduced in 1906, weekly hours were still 60 and annual 3 120 (no paid holidays) The estimate for 1970 is 2,100 hours/year [2]

U S. hours (Barger)	*1869*	*1879*	*1889*	*1899*	*1909*	*1919*	*1929*	*1939*
Agriculture	57 9	56 9	55 9	54 9	53 9	52 9	51 9	50 9
Mining	44	43	42	39	38	36	38	27
Manufacture	56	55	53	52	51	46	44	38
Commerce	66	66	66	65	59	56	54	48

National Product per head 1869 and earlier from Goldsmith (priv. comm)

[1] *Lifetime Allocation of Work and Leisure*, U S Government Printing Office 1968

[2] Rustant, *Analyse et Prevision*, Oct 1970

Monash University, Victoria, Australia

THE CONSUMPTION FUNCTION WHEN CAPITAL MARKETS ARE IMPERFECT: THE PERMANENT INCOME HYPOTHESIS RECONSIDERED[1]

By J. S. FLEMMING

I

THE permanent income hypothesis—that both individual and aggregate consumption is determined by the 'trend value' rather than the current level of income—is one of the most widely accepted post-war innovations in Economics Its proposer, Milton Friedman, also emphasizes that his theory implies a much lower value of the multiplier than was entertained by Keynes [2]

Friedman reached his conclusions by adopting an explicitly capital-theoretic approach to the consumption decision and applying the model so derived both to budget studies and to long runs of aggregate consumption and income data One of the assumptions of Friedman's model is that capital markets are perfect although he admits that 'the rate of interest at which an individual can borrow on the basis of his future earnings may . differ from the rate at which he can lend'[3] he chooses to 'neglect these complications'.[3]

In the next section of the paper Friedman's model of intertemporal choice is extended to conditions of capital market imperfection. The individual consumption function derived is shown to be substantially non-linear in current income if the elasticity of income expectation is low,[4] even for small divergences between lending and borrowing interest rates This implies that not only the *level* of permanent income but also the distribution of current income *relative* to permanent is relevant in the determination of aggregate consumption [5]

In particular it is shown that if the main deviations of current from permanent income are due to unemployment and the consequent interruption of some households' labour incomes, and if the capital market imperfection inhibits borrowing by the unemployed, a multiplier is

[1] I am grateful to M S Feldstein, P J Hammond, J R Hicks, M FitzGerald Scott, D W Soskice, G C Winston, and J F Wright for comments on previous drafts

[2] M Friedman, *A theory of consumption function*, NBER, Princeton, 1957, p 238

[3] Ibid, p 17.

[4] As postulated by Friedman and many Keynesians see, for example, A Leijonhufvud, *On Keynesian economics and the economics of Keynes*, New York, 1968

[5] Friedman, of course, recognized the potential relevance of income distribution effects (see, e g ibid, p 19) but did not consider the size distribution of income changes which is central in what follows

generated identical to that of Kahn's 1931[1] article which inspired Keynes's analysis.[2]

It is, however, possible that deviations of aggregate income from its 'permanent' level would not involve a small proportion of the work force becoming unemployed but rather a widespread reduction in working hours with a quite different impact on aggregate consumption spending. Section III presents some evidence supporting the hypothesis that employment bears the brunt of adjustment in the relevant short run and that income deviations are not widely distributed. It also discusses some evidence on the consumption response of the unemployed to their income loss. Both these topics raise questions about the length of Keynes's Marshallian 'short-run'. This issue is taken up in Section IV where a contrast is drawn with the notion of short run implicit in Friedman's permanent income hypothesis. Finally, Section V examines the attempts of Clower and Leijonhufvud to resuscitate Keynesian multiplier analysis without making capital market imperfection central to their analysis.

II

The permanent income hypothesis implies that consumption will exceed income if the latter drops sharply. It thus requires either that households have a 'cushion' of marketable assets or the capacity to borrow against future earnings. If neither of these is feasible there is clearly no alternative to reducing consumption. It is, however, less obvious that even a small interest differential will induce quite sharp consumption reductions in response to falling income.

The reason is simply that the marginal rate of substitution between current and future consumption is equal to $(1+r)$ where r is the rate of interest. Suppose that a consumer with no assets saves a little when income is at its permanent level: if income falls by more than his savings rate he can only maintain his consumption by dissaving, but this cannot be optimal. If the consumer dissaves r rises from the lenders' to the borrowers' rate: this must be accompanied by a drop in current consumption relative to its future level. In fact the following formalization of the argument shows that over a range of current incomes somewhat lower than permanent income the marginal propensity to consume is unity.

The argument is simplified if utility can be represented as a function of current consumption (C_0) and the endowment left for the next period (E_1),[3] thus

$$U = U(C_0, E_1)$$

[1] R. F. Kahn, 'The relation of home investment to employment', *E.J.*, vol. xli, no. 162 (June 1931), pp. 173–98.

[2] For Keynes's acknowledgement of his debt to Kahn see J. M. Keynes, *The general theory of employment, interest and money*, London, 1936, pp. 113, 115, 119, and 121.

[3] The utility function thus defined must be expected to vary over an individual's lifetime as E_1 stands as proxy for consumption over a shortening future. However, it is appropriate

There is a budget constraint which is some function (depending on th capital market conditions) of E_0, the initial endowment ($E_0 = A_0 + Y$ where A_0 is the initial holding of marketable assets and Y_0 is first perio income), and of H_1 the value at time 1 of future labour and other incom receipts. H_1 is fixed throughout, reflecting completely inelastic expecta tions thus

$$E_1 = H_1 + (A_0 + Y_0 - C_0)(1+r)$$
$$= H_1 + (E_0 - C_0)(1+r).$$

We want to examine the relationship between C_0 and Y_0 when utility i maximized subject to the different capital market constraints Fig 1(

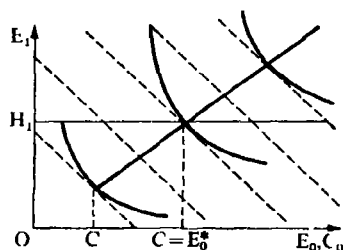
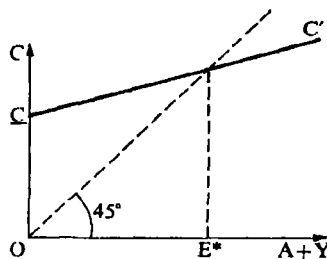



FIG. 1a FIG. 1b

represents the perfect capital market case in which the budget constrain is a straight line of slope $-(1+r)$, through the point (E_0, H_1) In this figur the level of consumption corresponding to a given level of the budge constraint is indicated by the abscissa of the point of tangency of th constraint with an indifference curve, while, as noted, the level of initia endowment is indicated by the abscissa of the point of intersection of th budget constraint with $H_1 H_1$ Consumption, C_0, can thus be drawn as a function of E_0 ($= A_0 + Y_0$) as in Fig 1b the consumption function $\underline{C}C'$ ha a slope less than unity as long as the expansion path (locus of tangencies in Fig. 1a is positively, and the budget constraint negatively, inclined

In Fig. 2a the budget line reflects a divergence between lending an borrowing rates of interest, the latter being the higher· this ensures tha it pays to sell assets earning the low rate before borrowing at the high rate The divergence of the two rates means that the single point, (E^*) in Fig 1a at which no financial assets are held ($C_0 = E_0$), becomes an interval ($-\underline{E}\bar{E}$ in Fig 2a; to this interval there corresponds a section of the consumptio function, ($C - C''$) in Fig 2b, over which the marginal propensity to con sume is unity [1]

for our short-run analysis and could also be justified in other contexts for a population wit a stable age structur

[1] The saving function implicit in Fig 2b is clearly very similar to that of Kaldor's 194 Trade Cycle model, *E J*, vol. l, no 197, pp 78 92 though he places as much emphasis o the redistribution of income from wages to profits as between the employed and unemploye the latter being central to the following analysis

Finally, if no borrowing at all is possible against H_1 the situation is as in Figs 3*a* and *b*, where the marginal propensity to consume is unity up to an endowment $\bar{E}$, beyond which it falls to the perfect capital market — permanent income—level

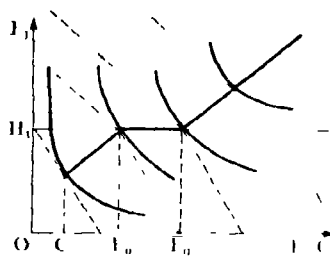



Fig. 2*a* Fig. 2*b*

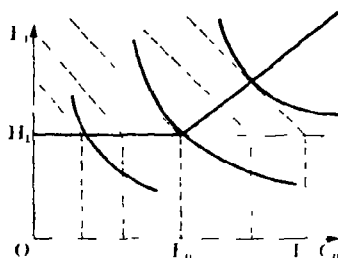
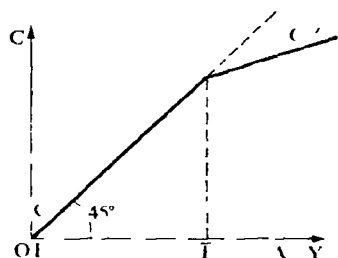



Fig. 3*a* Fig. 3*b*

The aggregate consumption functions implied by imperfect capital markets depend on the way in which the distribution of income deviations generate distribution effects. However it should be noted that the preceding analysis is essentially short-run in as much as H_1 is assumed invariant in a reasonably long time series H_1 will vary with the trend of *per capita* income shifting all the short-run functions—any distribution effects are to be seen as short-run phenomena. A simple permanent-income-type distributed-lag consumption function may well fit the data better than any *ad hoc* alternative not designed to reflect a specific hypothesis about the distribution of transitory income. Such a hypothesis is provided by the Keynesian employment adjustment assumption.

On getting a job a previously unemployed man adds to his consumption the amount

$$c(w)-c(u),$$

where $c(\ )$ is the relevant individual consumption function, w is the real wage, and u the real value of the receipts of the unemployed (Kahn's

'dole')[1] This additional consumption is to be compared with the add tional output which Kahn took to be equal to the real wage w On th assumption, and ignoring any consumption effects of changes in tl distribution of income between profits and wages consequent on change in w, the social marginal propensity to consume is

$$(c(w)-c(u))/w.$$

In the absurd limit when $u = 0$, and $c(u) = 0$,[2] we can see that th relevant social marginal propensity to consume will be more closely relate to the individual *average* propensity to consume of wage earners $(c(w)/u$ than to any mean of their marginal propensities $(dc(w)/dw)$

If those liable to unemployment have very few initial assets A_0, an if w is not much greater than $\bar{E}$ $c(w) \approx w$, while if u is greater than (which is zero in the third case considered above) $c(u) \approx u$ Thus once th unemployed have run their assets down to zero the Kahn/Keynes socia marginal propensity to consume would not differ significantly from

$$(w-u)/w = 1-(u/w)$$ [3]

The associated multiplier (w/u) might be estimated as the reciprocal of th ratio of the compensation received by the unemployed to the wage rat if one were prepared to ignore the assets (if any) of the unemployed Kahn took this ratio to be 'rather less than 1/2'[5] for the U K in 1931 which implies a marginal propensity 'rather more than 1/2' Is this th basis for Keynes's assertion that 'the marginal propensity to consum seems to be much nearer to unity than to zero'?[6]

III

Before considering the empirical standing of the Kahn/Keynes consump tion function interpreted in this way it should be emphasized that the concentration on the effects of individuals changing from being unemploye to being employed does not, in general, maximize the social margina propensity to consume (smpc) The model set out above implies that th

[1] If there is a public dole it might be argued that taxes should be brought into th picture this I refrain from doing for simplicity and coherence with the relevant literature Other justifications would be that the dole might be paid by creating money, or that al taxes fall on classes other than those liable to unemployment In the case of direct taxatio a Keynesian might be able to use this last argument, in conjunction with the consumptio function here defined, to rebut the argument that taxes lower the multiplier

[2] 'i e the unemployed live(d) on air', Kahn, p 189

[3] This is Kahn's equation (3) (p 184) with unity for his m'—the proportion of th increase that takes place in his income when he becomes employed devoted to hom produced consumption goods'

[4] This multiplier applies to expenditure it should be noted that transfers not directe to the unemployed will have an impact on expenditure as conventionally described

[5] Kahn, p 185

[6] Keynes, p 118

smpc would be unity if income changes were concentrated on people who, owing to short time working, had suffered a loss of income sufficient to raise their mpc to unity while not being drastic enough either to make them eligible for the dole or to make borrowing attractive

Thus when it is argued that the evidence suggests that employment adjustment is more important (in the relevant short run) than adjustment in hours of work it must be remembered that by the latter is meant general, economy-wide, adjustments of hours · short time working concentrated in particular industries or regions would typically generate a larger multiplier than the pure employment-adjustment multiplier attributed to Kahn and Keynes Evidence on the dissaving and borrowing opportunities is also considered below

In 1949 Modigliani wrote—in this context— that it 'is well known (that) cyclical fluctuations in income are much more the result of changes in unemployment than of changes in the income of the employed'[1] Modigliani refers only to evidence from the American slump, in what follows post-war studies of a number of countries are considered

For the United Kingdom the National Board of Prices and Incomes[2] concluded 'hours of work of operatives in manufacturing appear to respond [promptly] to cyclical and shorter term changes in demand . . changes in employment follow much later and are spread over the following year The movement of employment is more than twice that in hours ' (in proportional terms)

Contrary to the impression given by the Board's report a strikingly similar pattern emerges for the United States[3] Nadiri and Rosen have published the results of an econometric study in which employment, hours of work, investment, and capital utilization all respond in an interdependent manner to changes in demand[4] The relevant findings illustrated in Fig 4 can be summarized as follows

While hours respond more rapidly, in the sense that the peak of response is within one quarter, even in the first quarter the proportionate adjustment in numbers is almost twice that of hours (0 38 against 0 21) by the end of a year it is about seven times greater as the adjustment of men continues[5] while hours revert to normal.

[1] Franco Modigliani, 'Fluctuations in the saving income ratio', *Studies in Income and Wealth*, xi (New York NBER, 1949), p 387

[2] N B P I Report No 161, *Hours of work, overtime and shiftworking*, Cmnd 4554 (1970), Appendix C 'Relationship between hours of work, productivity and employment', p 114

[3] Ibid, p 111 where reference is made to Ulman's contribution (Ch VIII) to Caves (and associates), *Britain's economic prospects*, Brookings Washington and London, 1968 Ulman studied the responsiveness of *overtime* hours in various U K industries to G N P but this odd analysis made no allowance for the two big changes in *normal* hours in 1959/60 and 1964/5 which were substantially offset by increases in overtime

[4] I Nadiri and S Rosen, 'Interrelated factor demand functions', *A E R* (lix) 1969, pp 457-71

[5] Indeed it overshoots as, in the short run, men are substituted for the less adjustable capital their results suggest that employment adjusts much (four or five times) faster than capital

Unfortunately Nadiri and Rosen do not consider the possibility asymetrical responses to increases and decreases their data peri (1947–62) is dominated by t former while our interest is par cularly in the latter (inasmuch it is the cumulative nature of d pression that is to be explainec Moreover, they take no account any feed-back from factor deman to product demand although tl sector studied (total manufacturin contributes a significant proportic to total income Nevertheless the work offers some support, for wh it is worth, to the proposition th adjustment in men is more important, even in the 'short-run', than adjustment in hours, despite the 'quasi fixed' nature of the factor labour

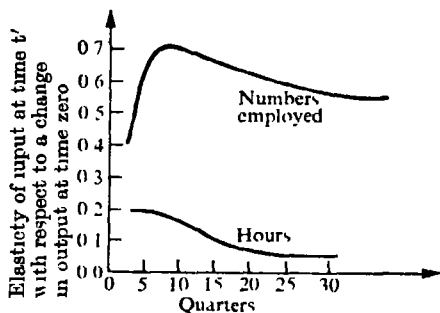



Fig 4

Brechling and O'Brien studied employment adjustment in twelve coun tries and although they do not use data on hours, investment, or capit utilization, it is shown in the footnote below that their results indicate th within a year the proportionate change in employment will substantiall exceed that in hours [2]

Turning now to the effect of unemployment on an individual's consum tion the following propositions seem relevant

(a) unemployment is concentrated—except in severe depression—o those least likely to have acquired marketable assets casual an unskilled workers, and the young who suffer either as school-leaver or from last-in-first-out firing rules,

(b) it is virtually impossible for such people to borrow against futur earnings however confident they may be of rapid re-employment,

(c) their liquid resources will not typically cover more than a very fev

[1] W Y Oi 'Labour as a quasi fixed factor', *J P E* 70 (1962), pp 538–55

[2] F P R Brechling and P O'Brien, 'Short run employment functions in manufacturing' *R E Stats* vol xlix, no 3 (1967), pp 277–87 This study of twelve countries (Austria Belgium, Canada, France, Germany, Ireland, Italy, Netherlands, Norway, Sweden, U K and U.S) suggests that, on average, 30 per cent of the divergence between desired employ ment (based on exogenous 'normal' hours-per-man) and actual employment is eliminate per quarter —50 per cent in six months The model assumes that the shortfall is made u by 'transitory' adjustment of hours On the assumption that the elasticity of output witl respect to employment and hours per-man is the same these figures indicate that after si months the proportionate employment and hours responses would be equal and th employment response predominate thereafter. In fact there are many reasons for expectin the elasticity of output with respect to hours per-man to exceed that with respect t employment in which case the previous conclusion is strengthened (See M S Feldstein 'Specification of the labour input in the aggregate production function', *R E Stud* xxxiv (1967), pp 375–85)

weeks' normal outgoings—including any contractual savings commitments they may have,

(b) uncertainty as to the duration of unemployment might then lead to skimping even though assets might cover normal outgoings for the expected duration,

(c) the existence of transaction costs, especially on household durables, will influence the timing of asset disposal and may inhibit it if it is anticipated that a recovery of fortune would shortly warrant their re-acquisition

These *a priori* assertions are supported by the work of P A Klein on 'Financial adjustments to unemployment'[1] in the U S. 1954–8 the unemployment rate was twice as high in the age group 20–4 as in the civilian labour force in 1956,[2] only 30 per cent of Klein's sample of 1,836 unemployed were initially debt-free as against 41 per cent for all spending units,[3] and 43 per cent of the unemployed had *no* liquid assets as against 26 per cent for all spending units [4] only 30 per cent had non-zero assets and zero debt [5] In a smaller sample of 319 unemployed from Pittsburgh for whom more detailed asset data were collected 36 per cent had non-zero assets and zero debt,[6] but only 20 per cent had (gross) liquid assets in excess of $500[7] which represented only seven weeks' *average* loss of income *after* allowing for unemployment compensation[8] (the average duration of unemployment was over 18 weeks) Borrowing, of which half took the form of non-payment of bills (including hire purchase and mortgage obligations), amounted to 11 per cent of the net loss of income [8]

This data leaves open the question of whether one should explain consumption in terms of those who have been unemployed for, say, more than six weeks, rather than unemployment *simpliciter* Leijonhufvud, referring to Klein's study, says that the unemployed 'seem to maintain their consumption standards for several months' [9] Klein does analyse his data by duration of unemployment but presents the results in such a way as to [su]ggest an exaggerated delay before the loss has a substantial impact on [co]nsumption

Klein divides his sample into five groups by length of unemployment: [for] each group he computes the average loss of income (net of unemploy[me]nt compensation) and the average reduction in consumption. the ratio [of] these is termed the '*marginal* propensity to consume' and, in his table [this] only exceeds 0·70 for those unemployed for more than 19 weeks [10]

[1] NBER, New York, 1965 [2] Ibid, Table A–1 [3] Ibid, Table A 5
[4] Ibid, Table A-6 [5] Ibid, Table 14, p 52 [6] Ibid, Table 11, p 47
[7] Ibid, Table 10, p. 45. [8] Ibid, p 3
[9] 'Keynes and the classics' (I E A London, 1969), p 44
[10] Klein, op cit, Table 2, p. 23

However, a simple adjustment of his table to allow for the much lower initial adjustment to consumption suggests that this level is reached within ten weeks, and quite probably earlier [1]

IV

At almost every turn in the preceding arguments we have come face to face with the question 'how long is Keynes's short-run?' The assumed factor adjustment mechanism and the subsequent consumption response of the unemployed requires that it be a period of, perhaps, three to six quarters rather than a mere one or two That Keynes's short-run admitted some labour adjustment goes without saying that he was referring to the 'wholly unemployed' when he used the words 'employment' and 'unemployment' is also obvious from the historical context of his work—even if 'unemployment' could be (and often is) interpreted merely as an excess supply of labour effort however distributed

Not only is it implicit in Keynes's use of period analysis (the application of comparative static techniques to dynamic processes) that the short period be long enough for equilibrium to be reached conditional on the initial conditions (as to capital stock, expectations, money wages) Keynes was very explicit about the short-run nature of his consumption function It is quite specifically set in the context of 'fluctuations in real income . which result from applying different quantities of employment to a given capital equipment' [2] Such a relation is no more that of the long aggregate time series from which the permanent income hypothesis draws its empirical support than that of the now discredited cross-sections with their apparent implication that redistribution was the answer to the threatened stagnation

Friedman's formulation, on the other hand, although it admits of a distinction between short- and long-run consumption functions, fails to provide an unambiguous interpretation of the former If one defines the short-run marginal propensity to consume in terms of the effect of changes in current income on consumption, it is likely to be sensitive to the time units chosen · for instance, the quarterly marginal propensity can safely be assumed to be lower than the annual one

It is important to notice that it is in the nature of both types of short-run function to shift over time. in the Keynesian case it shifts as capital is accumulated—just as would a Marshallian short-run supply function In this light it is remarkable that Kuznets's data revealing the long-run

[1] The adjustment involves computing the ratio of the excess of consumption reduction of each group over that of the preceding group to its excess income loss

[2] Keynes, p 114 Notice the use of the word 'employment' where 'labour' would now be more conventional

proportionality of consumption to income[1] caused such a stir. Quite apart from the sentence cited above some instability of the Keynesian real income-consumption function is implicit in his denomination of the relationship in terms of *wage-units*. If the wage-unit-deflated relationship is stable in the longer run the income consumption relation deflated by commodity prices must be expected to shift with the real wage— reflecting capital accumulation and technical change.

Friedman cites Keynes faithfully, including a reference to 'wage units' but, in common with many other post-war writers, fails to appreciate the significance of Keynes's choice of units.[2] Hicks alone recognized its substance as substituting employment for income in the determination of consumption—but preferred to use real income as conventionally defined.[3] Of course in the short-run the two real income concepts are related by the production function—to that extent the choice is arbitrary. In the long-run, however, matters are very different: if real wages were a stable fraction of income Keynes's 'wage-unit consumption function' would be stable in the long-run even on the permanent income hypothesis. However, Keynes's function was, as we have seen, explicitly restricted to the Marshallian short-run;[4] moreover, Keynes was more modest about its stability than Friedman would have us believe.[5]

It is noteworthy that earlier exponents of Keynesian trade cycle theory appreciated this; Kaldor, in particular, not only incorporated a sigmoid short-run savings function in his 1940 article (as noted above, p. 162), he also had savings as a function of 'activity', the whole function shifting with income. He thus incorporated all the relevant results of the model set out above as underlying the Kahn/Keynes multiplier.

The shifting of the savings function over time is also a feature of Duesenberry's model[6] which suggests, without much supporting argument, that the expectations and aspirations which determine consumption would

[1] S. Kuznets, *National product since 1869* and *National income: A survey of findings*, both N.B.E.R., 1946.

[2] Friedman, p. 3, citing Keynes, p. 96. Writing everything in money terms Keynes's consumption function is, when linearized,

$$C/W = A + BY/W \quad (A > 0,\ 0 < B < 1)$$

Deflated by the price level P instead of the money wage W this becomes

$$C/P = AW/P + BY/P$$

which shifts with the real wage W/P.

[3] J. R. Hicks, *A contribution to the theory of the trade cycle*, Oxford, 1950, p. 12.

[4] Keynes, p. 114 (cited above).

[5] Keynes described it as 'fairly stable' (p. 96) while Friedman writes 'Keynes took it for granted that current consumption expenditure is a *highly dependable and stable* function of current income' (p. 3) (emphasis added). See also Keynes's discussion of cyclical effects (p. 97) where he anticipates Duesenberry's analysis of the role of habit in the short run.

[6] J. S. Duesenberry, *Income saving and the theory of consumer behaviour*, Harvard, 1949.

be based on previous-peak income [1] Duesenberry refers explicitly to the incidence of unemployment and the unavailability of credit as explaining the patterns of dissaving in successive cross-section studies [2] When Friedman fitted alternative models to long time series he found that the Modigliani–Duesenberry previous-peak model appeared to out-perform the geometric-distributed-lag formulation of the permanent income hypothesis [3] The model of Section II might be regarded as providing a firmer rationalization of this result since an equation incorporating the deviation between current and previous-peak income includes a good proxy for unemployment, a point emphasized by Modigliani [4] If capital markets are imperfect—as Duesenberry assumes this is an important variable whether or not one accepts a crucial psychological role for previous-peak income

V

Even those who have tried to save Keynes from the neo-classical counter-revolution have tended to confuse this issue clearly the imperfect capital market model re-establishes just the sort of current income constraint required by Clower[5] and Leijonhufvud [6] Clower stated firmly that a current, disequilibrium, income constraint was essential to Keynes's analysis: 'Keynes either had a dual decision hypothesis (Clower's theory) at the back of his mind, or most of the *General Theory* is theoretical nonsense' [7] Yet he has himself subsequently contributed an elegant paper on the consumption function in a capital theoretic context with an effectively perfect capital market [8]

The problem of reconciling Clower's dual decision hypothesis—that actually achieved (probably disequilibrium) factor sales constrain effective consumer demand—with the permanent income hypothesis was raised by Professor Patinkin in his introduction to the discussion of Clower's paper at the Royaumont conference [9] Clower's reply is reported in the following terms:

> Even . with assets the dual decision hypothesis would be relevant since, unless one supposed that assets somehow get replaced without getting purchased, a chronic gap between desired and actual factor sales would sooner or later force all assets to

[1] Ibid, p 82 [2] Ibid, pp 78–81 [3] Friedman, pp 146–50

[4] Modigliani, p 391 Modigliani makes no references to dissaving or borrowing opportunities.

[5] R W Clower, 'The Keynesian counter-revolution' in *The theory of interest rates*, the proceedings of an International Economic Association Conference at Royaumont, France, 1962, edited by F H Hahn and F P R Brechling London, 1965, pp 103–25

[6] A Leijonhufvud, *On Keynesian economics and the economics of Keynes*, New York, 1968.

[7] Clower, p 120.

[8] R W Clower and M B. Johnson, 'Income, wealth and the theory of consumption' in *Value, capital and growth*, essays in honour of Sir John Hicks, edited by J N Wolfe, Edinburgh, 1968, pp 45 96

[9] Hahn and Brechling (eds), p 301

the zero level unless the gap was reflected instead in reduced demand for commodity flows.[1]

This argument is unconvincing. Friedman would presumably argue that the permanent income hypothesis implies that 'chronic' divergence between desired and actual factor sales will not emerge. The forces tending to restore the economy to equilibrium are adequate provided that sufficient assets exist for the *temporary* adjustment process. Clower has, in fact, argued in a circle by appealing to the implications of *chronic* disequilibrium when its occurrence is central to the debate. Clower's reply is only effective in the absence of *any* forces tending towards equilibrium.

Leijonhufvud reveals his ambivalence on this question in the following two quotations, both of which come from the same page of his book.[2]

(i) a household's '*effective demand* in other markets will be *constrained* by the *income* actually achieved. this is the crucial point' (Original emphasis.)

(ii) ' . . receipts from currently realized sales of household services *do not in themselves* constitute the budget-constraint on current household purchases. ' (Emphasis added.)

A footnote to this last passage refers one to Chapter IV on Liquidity Preference—and the wealth effect—for a discussion of 'the operative constraint'.

While it is clear that in a *non-tâtonnement* process actual (disequilibrium) income must be of *some* relevance there remains an issue of degree which is not faced by these authors. The nearest that we get to clarification in the book is the cryptic remark (in a footnote) that it would be 'incongruous' to make conventional 'perfect capital market' assumptions, since to accept that and its corollaries, 'would all but obviate the current income constraint here under discussion'.[3]

Leijonhufvud returns to these capital market imperfections in his lectures 'Keynes and the classics'[4] where he does face the issue of the magnitude of the multiplier explicitly referring to the quality of the empirical results of 'the modern theories of the consumption function' and the way in which inelastic income ('wealth') expectations '"fit in", most naturally, with our description of individual behavior in the labour market'.[5] He there describes 'the multiplier as an illiquidity phenomenon'[6] which reflects difficulties of borrowing against future labour earnings. However, borrowing against future earnings is only necessary on the part of individuals who experience income falls large or prolonged relative to their initial rate of

[1] Hahn and Brechling (eds), p. 308.
[2] Leijonhufvud, p. 56.
[3] Ibid., p. 57, footnote 15.
[4] Published by the Institute of Economic Affairs, London, 1969.
[5] Op. cit., p. 43.
[6] Ibid., pp. 42-4.

saving and their holdings of saleable assets. The sale of assets is not likely to be a satisfactory method of financing individual consumption if aggregate income has fallen by more than the equilibrium level of aggregate savings While such a fall clearly occurred during the slump it is most unusual for an initial autonomous demand reduction to be of this magnitude

Conclusion

The conclusion of this discussion is that the magnitude of the multiplier process depends on the interaction of three distinct factors (i) the elasticity (and confidence) of 'income' expectations, (ii) the feasibility (and cost) of borrowing against future earnings, and (iii) the distribution of income *changes* relative to savings and asset holdings

It is the contention of this paper that though Keynes did not explicitly deploy these three strands in his multiplier argument they can be so deployed as to substantiate his claims for the multiplier Friedman on the other hand, specifically admits the possibility of capital market imperfection yet relies exclusively on inelastic expectations (i) in his assertion that the multiplier is small Clower emphasizes that this makes a nonsense of Keynes but leaves it to Leijonhufvud to introduce capital market imperfections (ii) specifically if somewhat obliquely

Apart from reviewing these issues the purpose of this paper is to focus attention on the third element (iii) and to suggest that the hypothesis that employment (rather than hours) bears the brunt of labour input adjustment over the relevant time span not only completes the underpinnings of the Kahn/Keynes's multiplier but is also consistent with the available evidence

Nuffield College, Oxford

THE GAINS FROM TRADE IN AND OUT OF STEADY-STATE GROWTH[1]

By ALAN V. DEARDORFF

IN recent years growth theory and trade theory have converged in a flurry of articles which analyse growth of open economies and trade among growing economies. The first of these, in the tradition of modern growth theory, used mathematical techniques.[2] More recently, several authors have applied geometric techniques, familiar to trade theorists, to the problem.[3] Several of these authors have pointed out that growing economies may suffer a loss in steady-state *per capita* consumption as a result of opening to trade, though it has not always been clear how likely or how serious this result would be.

In this paper we use geometric techniques to analyse the effect of trade on *per capita* consumption in steady state and during the approach to steady state. The method used has the advantage that a single diagram yields results for a number of different cases without requiring difficult manipulations of the curves.

We will assume throughout that an economy saves and invests a constant fraction of its income. There need be nothing optimal, of course, about such proportional saving, and therefore the conclusions we draw about gains and losses from trade are only second-best results.

In Sections I and II we will define the model we will use and derive the geometric tools from which our results will follow.[4] In Section III we will discuss in detail the effect of a change in the terms of trade on steady-state *per capita* consumption, and show that a closed economy can suffer from free trade. In Section IV we look at trade between two countries with identical technologies but different propensities to save. In Section V we look at trade between countries with different technologies. Throughout Sections II to V we assume that the consumption good is always more capital-intensive than the investment good. In Section VI we look at what happens if this assumption is reversed. Finally, in Section VII we drop the assumption that economies are always in steady states and look at *per capita* consumption as the economy adjusts towards its steady state.

[1] This paper originated as part of the author's Ph.D. thesis submitted to Cornell University. The author has benefited from discussions with Jaroslav Vanek, Trent Bertrand, Edwin Burton, and Jay Levin and from extensive comments and suggestions provided by W. M. Corden. Thanks are due also to Robert M. Stern and the Research Seminar in International Economics at the University of Michigan.

[2] Oniki and Uzawa (1965), Bardhan (1965), and Stiglitz (1970).

[3] Johnson (1971), Vanek (1971), and Bertrand (1971).

[4] This paper had been written before the appearance of the articles by Stiglitz (1970), Johnson (1971), and Bertrand (1971). We will refer occasionally in footnotes to relationships between this paper and the works of those authors.

I. The closed economy

We consider a two-sector economy which can produce two goods a consumption good and an investment good To produce these it employs capital and labour subject to production functions which are neoclassical—that is, they are homogeneous of degree one and have positive but diminishing marginal products [1] These functions are assumed to be continuous and differentiable everywhere The labour force grows at a constant rate, n The capital stock depreciates at a constant rate, μ, but is augmented at each moment in time by the amount of investment Investment is always equal to savings which is in turn equal to a constant fraction, s, of national income

The economy is in a steady state when its capital stock is growing at the same rate, n, as the labour force Letting K be the capital stock, L be the labour force, and I be investment in units of the investment good, we have as our definition of a steady state that

$$\dot{K}/K = (I/K) - \mu = n, \tag{1}$$

where the dot over K indicates its derivative with respect to time Letting $k = K/L$ be the capital-labour ratio we can solve (1) to get

$$I/L = (\mu + n)k \tag{2}$$

as the equation for *per capita* investment in a steady state

Because we have assumed neoclassical production functions in both sectors, it follows that we can represent these functions as in Fig 1, where sectoral *per capita* output is shown as an increasing concave function of the sectoral capital-labour ratio alone Here k_i is the capital-labour ratio in the ith industry, L_i is the labour force employed in the ith industry, I is the output of the investment good, and C is the output of the consumption good Of course if the economy is completely specialized in production of either good then the corresponding curve gives *per capita* output for the whole economy, in units of that good, as a function of the economy's capital-labour ratio, k

For any given capital-labour ratio the economy can produce anywhere along a transformation curve such as that shown in Fig 2 as TT' The endpoints of this curve are given directly by the curves in Fig 1 The economy will choose to produce at a particular point on the transformation curve, such as P in Fig 2, only if the marginal rate of transformation between the two goods equals the ratio of their prices. That is, it will produce at P only if the price ratio is that given by the slope of the tangent, AB All of this is familiar from traditional trade theory

[1] In order to assure the existence of steady states, we assume the 'Inada conditions' that in both industries the marginal product of capital is infinite if the industry capital-labour ratio is zero and approaches zero as the capital-labour ratio becomes infinite

Let C be consumption If the economy is closed, then its consumption must equal its production of the consumption good, so that $C/L = \bar{C}/L$ If the economy is open and faces prices given by the slope of AB in Fig. 2, then it can consume anywhere along the line AB. Exactly where this consumption point will lie depends of course on the propensity to save, s

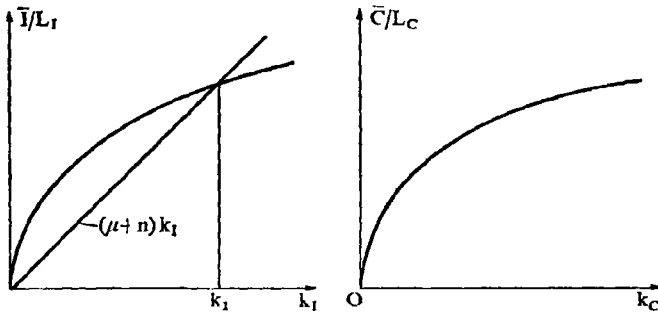



Fig. 1a, 1b

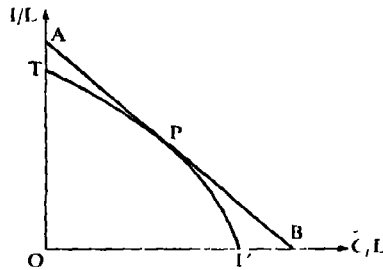



Fig. 2

Our object is to study *per capita* consumption, C/L, when the economy is in various steady states To do this we begin with a curve representing the level of *per capita* consumption that a closed economy could attain at each capital-labour ratio, assuming investment at each point to be just enough to keep the economy in steady-state growth This consumption possibility curve is the curve labelled $[C/L]$ in Fig 3, where the square brackets remind us that it refers to a closed economy We will explain its properties in a moment, but first a comment on its interpretation is in order

The $[C/L]$ curve does not represent a functional relationship between C/L and k The steady-state values of both of these variables are determined, though not necessarily uniquely, by the propensity to save as well as the other parameters of the model Thus the $[C/L]$ curve represents a locus of the steady-state values of both variables corresponding to all possible savings propensities

We can get a good deal of information about the nature of this locus by looking at the *per capita* production function for the investment good shown in Fig 1*a* It tells us the *per capita* production of the investment good that is possible if the economy completely specializes We have also drawn a ray from the origin, $(\mu+n)k_I$, which we know from equation 2 to be the amount of the investment good needed to maintain steady growth at each capital-labour ratio Comparing these two curves we see immediately that not all capital-labour ratios are consistent with steady growth If, as drawn in Fig 1*a*, the curves intersect at a positive capital-labour ratio, k_1, then

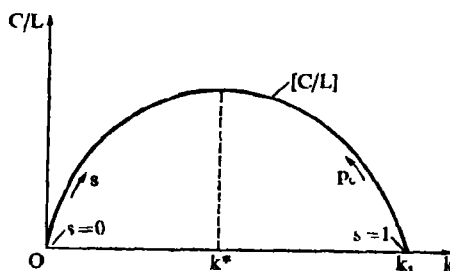



FIG 3

capital-labour ratios greater than k_1 cannot possibly be maintained To do so would require a greater amount of the investment good than the closed economy is capable of producing, even with complete specialization Furthermore, we are assured that the two curves will intersect in the manner shown if either the rate of population growth, n, or of depreciation, μ, is positive, and if the Inada conditions are satisfied [1] Thus steady states are possible only for capital-labour ratios between zero and k_1 inclusive, and the $[C/L]$ locus in Fig 3 takes on values only in this interval

We note also from Fig 1*a* that a steady state at $k = k_1$ requires complete specialization in the investment good Since the two production functions together rule out positive *per capita* production of any good if $k = 0$, we conclude that steady-state *per capita* consumption must be zero at both $k = 0$ and $k = k_1$ At capital-labour ratios between zero and k_1, however, complete specialization would yield more of the investment good than necessary to maintain steady growth There will therefore be resources left over for production of the consumption good, and steady-state consumption will be positive inside the interval Thus the closed economy's steady-state *per capita* consumption will be as drawn in Fig 3 zero at $k = 0$ and $k = k_1$, positive at $0 < k < k_1$, and undefined at $k > k_1$ [2]

[1] See footnote 1, p 174

[2] We have also drawn the curve as concave from below, a property which we will show later to be correct Concavity is suggested, of course, by the concavity of the production function in Fig 1*a*, if we identify the vertical distance between the two curves with the

The $[C/L]$ locus describes all of the steady states that a closed economy can attain. Which point on that locus the economy actually does attain will depend on its savings propensity. Furthermore, to different points on the locus there will normally correspond different price ratios needed to clear the markets for the investment and consumption goods.[1] It will be useful for our later analysis to know how the savings ratio and the price ratio vary along the curve.

The role of the savings ratio is easy to determine. If savings were zero, then no investment would take place and the steady-state capital-labour ratio would also have to be zero. If the savings ratio were unity, on the other hand, only the investment good would be produced, and we see from Fig. 1*a* that the steady-state capital-labour ratio would be k_1. Since our model is continuous, it follows that the savings propensity, s, that underlies the $[C/L]$ locus, must vary continuously from zero at $k = 0$ to unity at $k = k_1$. This relationship between k and s will also be monotonic, s increasing with k, if we make a further assumption, namely that the steady state corresponding to any given savings ratio is unique.[2] Making this assumption, we indicate with an arrow in Fig. 3 that s increases as we move to the right along the $[C/L]$ locus.

Finally, consider the behaviour of prices along the $[C/L]$ locus. As the capital-labour ratio rises, the price of the capital-intensive good will fall. This is necessary in order to maintain *per capita* production of the investment good that is proportional to k, as required in equation (2) for steady-state growth. Otherwise, if prices were to remain unchanged, the Rybczynski Theorem tells us that $\bar{I}/L$ would either fall absolutely or rise by a greater percentage than k, depending on the relative factor intensities.

We will assume, except in Section VI, that the investment good is labour-intensive. Thus in Fig. 3 we indicate with an arrow that the relative price of the consumption good, p_c, falls as we move to the right along $[C/L]$. The direction of this arrow will simply be reversed when we assume the investment good to be capital-intensive in Section VI.

The $[C/L]$ locus of Fig. 3, together with the indicated behaviour of s and p_c along it, provide us with as complete a picture of the steady states of the

resources left over for consumption after steady-state investment requirements are satisfied. This identification is correct, however, only if factor intensities in the two sectors happen to be identical.

[1] If factor intensities in the two industries happen to be identical then there will exist a unique price ratio consistent with incomplete specialization. In that case the equilibrium price would be the same at all points on the $[C/L]$ locus. If factor intensities are not identical, however, prices must vary along the locus.

[2] Uzawa (1963) has shown such uniqueness under the assumption that production of the investment good is relatively labour-intensive. Drandakis (1963) has shown that steady states will be unique regardless of factor intensities if the elasticities of factor substitution in the production functions are sufficiently large. Thus either of these assumptions suffices to give us monotonic variation of s along the $[C/L]$ locus.

closed economy as we need for our analysis It will be convenient, however before we go on to the open economy, to identify one further property of the closed economy. Since $[C/L]$ is zero at $k = 0$ and $k = k_1$ and is positive in between, it follows that it must reach a maximum at some $k = k^*$ in that interval [1] The savings ratio, s^*, to which this k^* corresponds is the familiar golden rule savings ratio. Therefore we will call k^* the golden rule capital-labour ratio and the corresponding p_c^* the golden rule terms of trade.

II. The open economy

We now consider an open economy that has some given savings ratio, but which may face any of a variety of international prices To each price that such a country faces, there will correspond a steady-state capital-labour

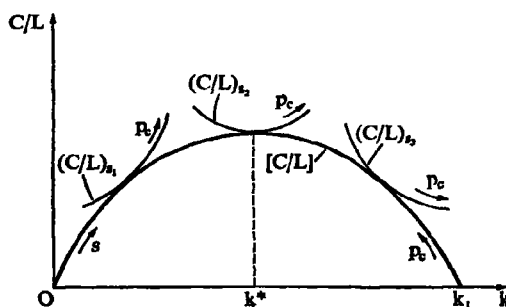



Fig 4

ratio and some amount of consumption per head in that steady state Thus we can imagine a locus, $(C/L)_s$, of *per capita* consumption for all steady-state capital-labour ratios attainable with the given savings ratio but with different prices, p_c. We have drawn three such loci in Fig. 4, each corresponding to a different savings ratio We must now show that they are drawn correctly

First we note that any $(C/L)_s$ locus must coincide at one and only one point with the $[C/L]$ locus of Fig 3 This point of coincidence would be reached if the world price equalled the price that would prevail in steady state in a closed economy with the same savings ratio. It is therefore the point of no trade along the $(C/L)_s$ locus Its existence and uniqueness

[1] k^* can be identified, as in one sector growth models, by the equality of the marginal product of capital with $(\mu+n)$. This marginal product must, of course, be measured in units of the investment good in order for it to be of the same dimension as $(\mu+n)$ To verify the criterion we merely assume the contrary If the marginal product of capital were greater than $(\mu+n)$ then only part of an additional unit of capital would have to be employed in the investment good industry to maintain steady growth at the now higher capital-labour ratio That part of the additional capital which is left over can be used to increase *per capita* consumption. A similar argument can be used to show that if the marginal product of capital is less than $(\mu+n)$, a small decrease in the capital-labour ratio will increase steady-state *per capita* consumption

follow from the continuous monotonic variation of s from zero to one along $[C/L]$.

Next we note that a $(C/L)_s$ locus cannot pass below $[C/L]$. This follows directly from the familiar static gains from trade. We know that trade permits a combination of consumption and investment outside of a country's transformation curve. In general, of course, this gain from trade need not imply an increase in consumption, since the gain could be absorbed entirely into increased use of the investment good. However, the loci $(C/L)_s$ and $[C/L]$ both represent steady-state solutions. This means that at any given capital-labour ratio, *per capita* investment must be the same on both curves. It follows that the static gain from trade must be devoted entirely to an increase in *per capita* consumption in the open economy over what it would be in a closed economy with the same steady-state capital-labour ratio. Notice, however, that this does not mean that an economy with a fixed savings ratio must gain from trade. We have compared open and closed economies with identical steady-state capital-labour ratios, and they must therefore have different savings ratios if the open economy actually trades. But we have established the desired result that the $(C/L)_s$ locus lies above $[C/L]$ except at the unique point where they coincide.

Finally, we note that at any strictly positive $[C/L]$ the closed economy must be incompletely specialized. It should be clear from the nature of our model that its solution must be continuous and smooth except possibly at points verging on complete specialization. Thus the $(C/L)_s$ locus must be smooth where it touches the $[C/L]$ locus, and since the two cannot cross they must be tangent. This establishes the essential features of the $(C/L)_s$ loci as drawn in Fig. 4.

Jaroslav Vanek has shown that when the price of the consumption good rises, an open economy like the one we are discussing will move to a higher steady-state capital-labour ratio. This is easily understood. An increase in p_c raises national income in units of the investment good for any given capital-labour ratio. Since a constant fraction of this income is saved, investment must increase and so must the steady-state capital-labour ratio. From this it follows that p_c must be rising as we move to the right along any $(C/L)_s$ locus. This we have again indicated with arrows in Fig. 4.

There is one more locus that will be of some use to us later on. That is the locus of steady-state *per capita* consumption for an open economy with fixed terms of trade but variable savings ratio.

Two such loci denoted $(C/L)_p$, have been drawn in Fig. 5.[1] Each is shown as a straight line in the neighbourhood of the autarky point where it touches

[1] These loci correspond exactly to the open economy investment requirements curves which Johnson (1971) labelled PQ in his article and which Bertrand (1971) corrected and further analysed. Ours are drawn as functions of K/L instead of I/L, but the transformation is straightforward, since in steady state K/L and I/L are related by equation (2).

$[C/L]$. This is a result of the homogeneity of the production functic According to Euler's Theorem, the value of each sector's output equals value of inputs used in the sector. The value of the country's output, th must be the value of the country's factor supplies, K and L. In *per cap* terms

$$Y/L = r(K/L)+w,$$

where r and w are the prices of capital and labour respectively and Y is t value of national output. From factor-price equalization theory we kn

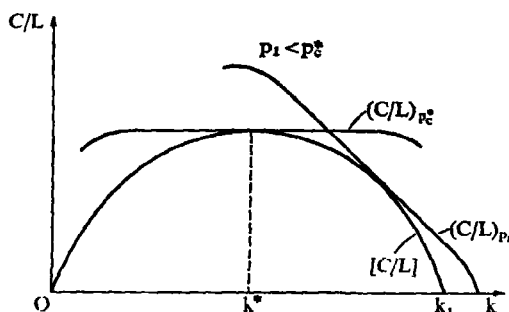



FIG. 5

that constant commodity prices imply constant factor prices as long as bot goods are produced. Thus in equation (3), *per capita* income is a linea function of the capital-labour ratio in a region of nonspecialization. Sinc the *per capita* investment needed for steady growth is also a linear functior of k, it follows that the difference between the two—steady-state consump tion per head—must be a linear function of k in the neighbourhood of the autarky point

Incidentally, the same argument we used to show that no $(C/L)_s$ locus can pass below $[C/L]$ applies as well to $(C/L)_p$ The gains from trade can influence only consumption when we compare two economies in steady state at the same capital-labour ratio Thus, it follows that the $[C/L]$ locus can never pass above any straight line drawn tangent to it at any point. This means that it must be concave—a fact which we have always considered when drawing it but have not, until now, proved.

III. The possibility of gains and losses from trade

We are now ready to assert our fundamental result regarding gains from trade. Consider any closed economy in steady state. Its capital-labour ratio and *per capita* consumption must be those indicated by the point on the $[C/L]$ locus corresponding to its savings ratio. At that point there is a $(C/L)_s$ locus which is tangent and which gives us the *per capita* consumption it can reach if it opens to trade at various prices. It is immediately

ɔvious from Fig. 4 that there will almost always be some prices at which ·ade would be harmful.

Suppose, for example, that the economy's savings ratio is s_3 in Fig 4. hen a small increase in p_c must cause its steady-state capital-labour ratio ◦ increase and its *per capita* consumption to fall This follows because at $[C/L]$ is downward sloping and $(C/L)_{s3}$, being tangent to it and smooth, ust also be downward sloping Similarly, if s_1 is its savings ratio, small ecreases in p_c will cause a fall in *per capita* consumption Indeed, ly when $s_2 = s^*$ is *per capita* consumption already at a minimum in utarky.

Next we notice that when a closed economy enters into trade at a higher rice for the consumption good, it must, in trade, be exporting the consumption good and importing the investment good Combining this fact ith the observations in the preceding paragraph, we can state the llowing theorem.

Theorem If an economy's savings ratio is equal to the golden rule savings ratio s^* which maximizes steady-state *per capita* consumption when the economy is closed, then any trade will increase its steady-state *per capita* consumption If $s < s^*$, then a little trade is worse than no trade at all if it involves exporting the investment good If $s > s^*$, a little trade is worse than no trade at all if it involves exporting the consumption good

The words, 'a little trade', in the above are essential In the previous ection we considered properties of the $(C/L)_s$ locus only in the neighbourɔod of its tangency with $[C/L]$ As the difference between the trade price d the autarky price becomes large we leave the neighbourhood of such tangency and the curve is quite likely to turn back up [1] Thus, our theorem valid only for small changes in the price from autarky [2]

The theorem is the central result of this paper In the following sections e will merely ask under what circumstances free trade might be expected, cording to this theorem, to be detrimental Thus, it is desirable, before e go on, to be sure that we understand why the theorem is true

A good clue to what is happening here is the important role played by e golden rule savings ratio, s^* As explained earlier, the golden rule

[1] In fact it must turn back up When the economy is producing only the investment ɔod, a further fall in p_c has no effect on the steady-state capital-labour ratio and can only crease *per capita* consumption, when the economy is producing only the consumption good further rise in p_c can affect C/L only by changing K/L, and K/L must rise with an increase p_c, thereby increasing C/L

[2] Our theorem agrees completely with Johnson's (1971) conclusions regarding *per capita* nsumption, so long as one looks only at savings less than the golden rule His failure to nsider what happens when savings is greater than the golden rule has also been pointed ıt by Bertrand (1971).

capital-labour ratio in the closed economy is characterized by the equa of the marginal product of capital with the investment per unit of cap needed to maintain steady growth [1] Thus, if a closed economy starts steady state below k^*, the marginal product of capital must be greater t the investment necessary to maintain a unit of it. If the economy were t to export the investment good, it would lose more in productivity tha would gain by not having to maintain the investment good as capi Alternatively, at $k > k^*$, the marginal product of capital is too small that imports of the investment good add less in productivity than is nee to maintain the increased capital-labour ratio

We see, then, that a country will incur a decrease in steady-state *capita* consumption whenever it moves away from the golden-rule capit labour ratio This is true regardless of whether the movement results fr a change in the economy's savings ratio or from a change from autarky trade In the latter case, however, the conclusion holds only if the volu of trade is small For as the volume of trade becomes large, the usual sta gains from trade become significant and outweigh the loss due to the no optimal capital-labour ratio

IV. Trade generated by different savings propensities

We now consider two countries with identical technologies but differe savings ratios, and attempt to determine how steady-state *per capita* co sumption in the two countries will be affected by trade between them. I order for steady states to exist under free trade it is necessary that w assume the rates of population growth to be the same in both countries With these assumptions the various loci that we derived in Sections I an II relating steady-state C/L to the capital-labour ratio will be identica for the two countries.

We denote the countries by A and B, and their savings ratios by s_a and s_b with $s_b > s_a$ We can determine the effect on steady-state C/L of the opening of trade by first identifying on the $[C/L]$ curve the two closec economy steady states and the corresponding steady-state prices Noting the arrows connected with $[C/L]$ in Fig 4 we see that steady-state p_c will be lower in country B than in country A since $s_b > s_a$ Therefore, free trade will cause a fall in p_c for country A and rise in p_c for country B. If these changes are small enough, our theorem tells us that the effect on steady-state C/L in each country will depend on whether their respective savings

[1] See footnote 1, p. 178.

[2] See Vanek (1971) Actually, as he argues there, if the population growth rates differ in the two countries the faster growing country will eventually become so large relative to the other that it may be regarded as closed, while the slower growing country may eventually be regarded as facing terms of trade given from outside

ratios are smaller than, equal to, or greater than the golden rule savings ratio, s^*. Since p_c falls for country A it will move to the left along its $(C/L)_{s_a}$ locus If $s_a < s^*$, then this movement implies a fall in steady-state C/L, while if $s_a \geq s^*$ it implies a rise. Similarly, p_c rises for country B, requiring movement to the right along its $(C/L)_{s_b}$ locus Thus, in country B, C/L will rise if $s_b \leq s^*$ and fall if $s_b > s^*$

If s_a and s_b lie on the same side of s^* (and if they are close together so that price changes will be small enough to apply the theorem) then the country whose savings ratio is closest to s^* will gain from trade while the other will lose Of course, if the two savings ratios are far apart, then the change in price from autarky to free trade may be large and both countries may gain from trade

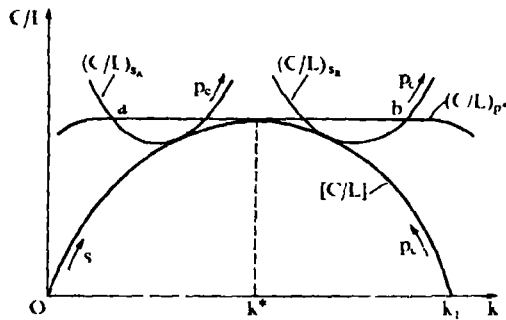



FIG. 6

If the savings ratios lie on opposite sides of s^* $(s_a < s^* < s_b)$, then if our theorem applies it will mean a fall in C/L for *both* countries It is not clear, however, that in this case the price changes can be assumed small enough to apply the theorem To see this we employ the $(C/L)_p$ locus discussed at the end of Section II. In Fig 6 we have drawn two appropriate $(C/L)_s$ loci, one on either side of s^*, and in addition have added the $(C/L)_{p^*_c}$ locus of Fig. 5 It should be clear that if, with the opening of trade, p_c should happen to settle precisely at p^*_c, then country A must move to the point a and country B must move to the point b Thus, both countries will have increased C/L through trade If the free trade p_c differs from p^*_c, then one country will end up above the $(C/L)_{p^*_c}$ locus and one country will end up below it. In any case, as long as both curves intersect $(C/L)_{p^*_c}$ in its linear segment, it is impossible for both to lose from trade.

The linear segment of any $(C/L)_p$ locus corresponds, as we have seen before, to incomplete specialization Outside of this region both ends of the locus curve toward the k-axis. Thus if two countries have savings ratios that are sufficiently far apart, both may completely specialize with trade

operating on the curved portions of a $(C/L)_p$ locus. This suggests t possibility that steady-state *per capita* consumption could be lower w trade than without for both countries. Unfortunately most of our inforn tion about both the $(C/L)_p$ and the $(C/L)_s$ curves is valid only with inco plete specialization, since it was based on continuity and smoothness functions in the neighbourhoods of autarky points. Thus we are n equipped either to prove or disprove this possibility that both countr might lose from trade

We can, however, give intuitive plausibility to such an occurrence neither country completely specializes in free trade, then factor prices w be equalized between them and world production will be efficient If, (the other hand, either country does specialize, then factor prices will rema unequal in free trade. This means that world production is inefficient a compared to a world with perfect factor mobility. Now in a static contex this inefficiency cannot negate the gains from trade, since it must be con pared to the even greater inefficiency of autarky But in the dynamic mod we consider here, free trade leads the capital-labour ratios of the two cour tries to move further apart than they were in autarky, thus increasing th production inefficiency that results from complete specialization.

We have been interested here in gains from trade However, our analysi also demonstrates clearly the effect of trade on capital accumulation Manipulation of Fig. 4 shows us that with the opening of trade the steady state capital-labour ratios in the two countries must move in opposit directions and become further apart than they were under autarky. Th reason for this result is straightforward. The country with the highe savings propensity attains a higher capital-labour ratio in steady state as a closed economy. This gives it a comparative advantage in the capital-intensive consumption good. Trade, therefore, decreases the price of the investment good in the high-saving country and increases it in the low-saving country. This causes investment—and therefore steady-state capital-labour ratios—to rise in the former and fall in the latter.

This, in a sense, is an extension of familiar comparative advantage results. A country which has a comparative advantage in production of one good will, with the opening of trade, shift its production towards that good Similarly, a country which, because of its higher propensity to save, has an advantage in the accumulation of capital, will, with the opening of trade, be able to support an even higher ratio of capital to labour We will see in Section VI, however, that this result holds only if the consumption good is relatively capital-intensive If the reverse is true, then capital-labour ratios will be drawn closer together by trade, since a high savings rate will now generate comparative advantage in the investment good instead of the consumption good.

V. Trade generated by different technologies

We have just seen how the gains from trade will behave if that trade is generated by differences in propensities to save We look now at the same problem when trade results from differences in technologies

Suppose that two countries, A and B, have identical savings ratios and that the technology is Hicks-neutral more efficient by the same percentage in both industries in country B than in country A. It is not hard to see that the steady-state autarky price of the consumption good will be lower in B than in A as long as the consumption good is relatively capital-intensive To see this, suppose that initially both countries have the same capital-labour ratio. B's transformation curve will then be a uniform radial expansion of A's Since their savings ratios are identical they will both find short-run equilibrium at the same price, with B producing more of both goods than A. If A happens to be in its steady state, then B must necessarily be growing, since their steady-state capital requirements, $(\mu+n)k$, are the same This shows that the autarky steady-state capital-labour ratio must be higher for B than for A. We need then only use the familiar result that an increase in the capital-labour ratio lowers the relative price of the capital-intensive good to conclude that the autarky steady-state p_c must be lower in B than in A.

We now know that free trade will cause an increase in p_c for the more efficient country, B, and a decrease for A Using our theorem it then follows that if the difference in technologies is small—so that the price change is small—then the more efficient country, B, can lose from trade if its savings ratio is above s_b^*, while country A can lose from trade if its savings ratio is below s_a^*, where s_a^* and s_b^* are the golden rule savings ratios corresponding to the two different technologies.

A second possibility that is easily analysed is that of a Hicks-neutral technical advantage in only the consumption goods industry In that case, for a given capital-labour ratio, the more efficient country's transformation curve will be a horizontal percentage expansion of the other country's transformation curve, assuming the consumption good to be measured on the horizontal axis. It is easily verified that the autarky steady-state capital-labour ratios will be identical in the two countries, but the p_c will be lower in the more efficient country Thus, our application of the theorem in the previous case continues to hold

If there is Hicks-neutral technical advantage in only the investment goods industry, then no general statement can be made The net effect will be a combination of the two just mentioned a uniform technical advantage in both industries plus a technical disadvantage of the same percentage in only the consumption good. Since the first of these means

a lower p_c and the second means a higher p_c, the net difference betwee autarky prices in the two countries will depend on which effect is stronge

VI. The case of the capital-intensive investment good

We have assumed, since the end of Section I, that production of th consumption good is relatively capital-intensive compared with productio of the investment good This is the familiar capital-intensity conditio used by Uzawa to assure stability and uniqueness of the steady state fo a closed economy with a given savings ratio It is not, however, a necessar condition for stability and uniqueness.[1] Nor is there any reasonabl economic justification for assuming that it is satisfied Therefore, in thi section we will outline the changes that must be made in our analysis i we make the reverse capital intensity assumption [2]

Actually, very few changes are necessary We observed in Section I that a capital-intensive investment good causes the price of the consumption good, p_c, to rise rather than fall as we move to the right along the $[C/L]$ locus All other properties of the $[C/L]$ locus are maintained, including the movement of s along $[C/L]$, since we continue to assume that a unique steady state is associated with each savings ratio. Furthermore, the arguments used to derive the $(C/L)_s$ loci and their properties were independent of the capital-intensity assumption [3]

We see then that the only difference between the picture in Fig 4 and the picture we need for our new capital-intensity assumption is that the direction of the arrow indicating price change along $[C/L]$ is reversed. It is easily seen that our theorem as stated in Section III continues to be true It did not rely on the direction of price change along $[C/L]$.

Our application of the theorem in Section IV does, however, require revision Now if two closed economies are in steady state, the one with the higher savings ratio will also have a higher p_c, and thus will experience a fall in p_c when trade is opened between them It is easily verified that if the savings ratios of the two countries are on the same side of the golden rule s^*, then the one closer to s^* will lose from trade while

[1] See footnote 2, p 177

[2] For completeness we should, perhaps, mention what happens if factor intensities in the two sectors are identical As mentioned in footnote 1, p. 177, there would then be only one price ratio consistent with incomplete specialization, regardless of the capital-labour ratio. It follows that operation anywhere above the $[C/L]$ locus with trade requires complete specialization. Each $(C/L)_s$ locus then has a kink where it touches $[C/L]$, and our arguments which rely on the smoothness of these loci do not work In fact it appears that the $(C/L)_s$ loci are essentially V-shaped, rather than U-shaped, and any trade is beneficial. This case is rather pathological, however, since the two production functions are presumably independent.

[3] Actually, it turns out that while the $(C/L)_s$ loci must be convex from below when C is K-intensive, they may be concave from below when I is K-intensive, at least in the neighbourhood of their tangencies with $[C/L]$ However, since our results do not depend on the curvature of these loci, we can ignore this technicality.

the one further from s^* will gain from trade If, on the other hand, the savings ratios are on opposite sides of s^*, then both must necessarily gain from trade.

VII. The adjustment towards steady state

We have occupied ourselves so far in this paper entirely with steady states Unfortunately, steady states are remote at best To discuss gains and losses from trade that will not be realized for decades or even centuries may seem a waste of effort In recognition of this quite valid objection we look briefly in this section at the behaviour of the model when it is not in a steady state We restrict attention here to a single small economy which, if it trades, faces a terms of trade that is given and fixed through time

(a) *Initially in steady state*

Suppose that initially the economy is closed and in its appropriate steady state as given by a point on the $[C/L]$ locus Then suppose that it opens itself to trade with a world in which the international price p_c is higher than its domestic price It will of course begin to export the consumption good. But more important for our purposes is the fact that its *per capita* consumption will fall The reason for this is that its output as a closed economy must have contained some of the investment good whose price has fallen Its national income in terms of the consumption good has therefore also fallen, and consumption is a constant fraction of that national income This result holds even after the economy has shifted resources into production of the consumption good, as may easily be verified with a transformation curve and several price lines

The opposite result holds if p_c falls with trade That is, *per capita* consumption will rise initially

These results may be added to the diagram in Fig 4, and we have done so in Fig 7. There we have drawn, for each of three savings ratios, the paths of *per capita* consumption as the economy adjusts to prices higher and lower than autarky. In every case the initial effect of the price change and the long-term effect of the change in the capital-labour ratio act in opposite directions If the long-run effect is not sufficient to offset the initial effect, then the gain or loss from trade is unambiguous For example, with $s = s_3$ and a fall in p_c, *per capita* consumption rises immediately above its autarky value and stays above it Similarly, with $s = s_3$ and a rise in p_c, C/L is ever after lower than it was in autarky In either case it turns out that our decision as to gains from trade looking only at steady states was correct the short-run effects only strengthen the conclusion

On the other hand, if the long-run effect more than offsets the initial effect, then whether or not there have been gains from trade depends on

the time horizon one wishes to use. When $s = s_1$ and p_c rises, then thei is initially a fall in C/L but over time it rises again and eventually surpasse its autarky value. To say that there has been a gain from trade is to tak a very long-run view of the welfare of the economy Still, if we take th short-run view, it is not clear that *per capita* consumption is the appropriat index of welfare—at least if saving is taking place voluntarily [1]

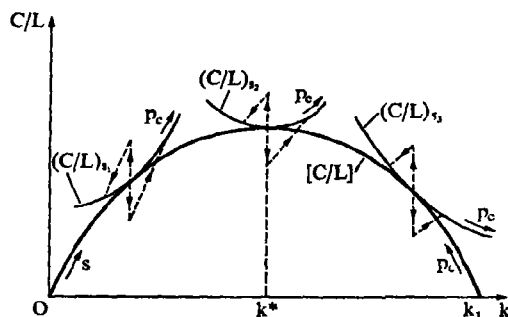



FIG 7

(*b*) *Not initially in steady state*

Before we leave this section we relax the assumption that the closed economy is initially in a steady state In its place we assume that the world price is identical to that which would prevail for the closed economy in steady state In these circumstances trade can have an effect only during the approach to steady state, but not in the steady state itself We then ask under what circumstances it would be desirable to open the economy to trade during that approach

Consider, for example, a closed economy with savings ratio $\bar{s}$. Let $\bar{k}$ be the steady-state capital-labour ratio associated with $\bar{s}$ for the closed economy, and $\bar{p}_c$ be the corresponding steady-state price of the consumption good. Suppose, however, that the economy is not initially in steady state, but that it starts with a capital-labour ratio, k^0, less than $\bar{k}$ If the consumption good is capital intensive, then the fact that k^0 is below $\bar{k}$ implies that p_c in the closed economy must initially be higher than the world price. This means that if the economy were to open to trade, the price of the

[1] Stiglitz (1970), using a model similar to ours but with different savings assumptions, found the possibility that steady-state *per capita* consumption would fall with trade, but he commented 'Should we, therefore, conclude that the opening of trade may have lowered social welfare? Obviously not' 'What has happened is that the opening of trade has changed the pattern of consumption over time and more consumption may be taken earlier . ' His comment applies as well to our model if savings is less than the golden rule But if savings is greater than the golden rule we see that a loss in steady-state consumption follows an even greater loss in consumption in the short run Stiglitz did not get this result because his savings assumptions prevented savings from ever being greater than the golden rule

›nsumption good would fall and *per capita* consumption would rise imme-
iately. This immediate gain from trade extends—in a sense—to the entire
›riod of adjustment, since p_c would remain above $\bar{p}_c$ for all time if the
›onomy remained closed, approaching $\bar{p}_c$ from above as the closed economy
›proached steady state By opening to trade, then, the economy will enjoy
›eater *per capita* consumption at each capital-labour ratio than it would
ave if it were to remain closed.

This does not mean, however, that the economy necessarily gains from
ade For the price change which increases *per capita* consumption also
duces *per capita* investment. Thus, while the economy's capital-labour
tio must rise towards the same steady state regardless of whether it is
›en or closed, it will approach that steady state more slowly if it is open.
nce *per capita* consumption increases with the capital-labour ratio, the
itial gain from trade may be offset by the faster growth of the capital-
bour ratio under autarky Thus we must compare a short-run gain from
ade with a possible long-run loss The desirability of trade during the
›proach to steady state must then depend on the country's attitude
›ward time preference

In the case just considered the economy begins its adjustment from a
ıpital-labour ratio below the steady state, $\bar{k}$ If instead k^0 is initially
›ove $\bar{k}$, then the results are reversed With $k^0 > \bar{k}$, and the consumption
›od capital intensive, the world price, $\bar{p}_c$, must be above the initial autarky
ice Opening to trade will then cause an initial loss of *per capita* con-
mption At the same time, however, *per capita* investment will increase,
› that the economy's capital-labour ratio will fall towards $\bar{k}$ more slowly
ıan it would if it remained closed. Thus the initial loss from trade might
›ssibly be offset after some time by a higher capital-labour ratio

If the investment good is capital intensive, then all of these conclusions
e again reversed The price of the consumption good will then be below
e world price when k^0 is less than $\bar{k}$ and above the world price when k^0
above $\bar{k}$. For all cases, however, we can say the following

> If opening to trade involves a fall (rise) in the price of the consumption good, then the economy will gain (lose) from trade initially; this immediate effect of trade may be offset later in the adjustment process, however, since the open economy's capital-labour ratio will be less than (greater than) what it would have been had it remained closed

It should be noted, in comparing the paths followed by the closed and
pen economies, that the capital-labour ratios of the two can differ at each
oint in time even though they approach the same steady state, $\bar{k}$ Of
›urse, this difference must itself approach zero, but it will never actually
e zero

VIII. Conclusion

We have devoted most of this paper to showing various ways that a growing economy may lose from trade. These results may seem to contradict the traditional static gains from trade analysis, but actually they do not. It is still true that an economy *can* gain from trade if it behaves optimally. As Trent Bertrand (1971) has shown, the locus of steady-state consumption possibilities with free trade dominates that of the closed economy.[1] By suitably adjusting the savings propensity the economy will never lose from free trade and will usually gain. We conclude, therefore with an attempt to identify which of the assumptions of our model were responsible for the loss from trade that we derived. The most obvious of these is the constant savings propensity, but we defer its consideration for a moment.

First consider the many other assumptions used in our model. These included constant exponential growth of the labour force, depreciation which is a constant fraction of the capital stock, and neoclassical production functions which are subject to the Inada derivative conditions. These or similar assumptions are needed in order that steady states may exist. For example, if population does not grow and if capital does not depreciate, then the capital-labour ratio would grow without bound.[2] There would be no steady state and questions concerning steady-state *per capita* consumption would be meaningless.

Our model also follows the tradition of neoclassical growth models in assuming that markets work perfectly. Specifically, we assume full employment of factors. Since our purpose has been to show that an economy may lose from trade, the fact that we have done so without invoking such market imperfections is probably an advantage.

We turn, then, to the savings assumption. As mentioned earlier, trade will never decrease steady-state *per capita* consumption if the economy saves appropriately. Our assumption that savings is a constant fraction of income is therefore the major reason for our result that the economy can lose from trade. But the same would be true for any savings assumption that is not based on individual maximizing behaviour. If, for example, savings were a constant fraction of profits rather than of income, it would still be true that steady-state *per capita* consumption could fall with trade.[3] Thus our proportional savings assumption is by no means the only savings assumption that can generate a steady-state loss from trade.

[1] These loci correspond in our analysis to $(C/L)_p$ and $[C/L]$ respectively.

[2] This could also happen if the Inada conditions fail to hold, or if depreciation, though positive, fails to increase in proportion to the capital stock.

[3] This has been shown by Stiglitz (1970) and is true only if the fraction of profits saved is less than one.

Our reason for assuming proportional savings is rather that it simplifies the analysis and that it permits a greater variety of autarkic steady states. The assumption that only profits are saved, for example, leads to a discontinuity in steady-state behaviour that makes the kind of analysis we used impossible And in addition, if only profits are saved, the closed steady-state capital-labour ratio cannot exceed the golden rule capital-labour ratio, k^* Thus all of the solutions associated with an economy which over-saves are ruled out from the start

Thus our proportional savings assumption has allowed us to demonstrate quite simply the possibility of a steady-state loss from trade Such a loss is likely to be possible with any savings assumption that is not inherently optimal, though to prove this would require a much more difficult analysis than we attempted here. In any case, the constant savings propensity does occupy a prominent place in economic literature—especially in growth theory—and its implications deserve to be explored

University of Michigan, Ann Arbor

REFERENCES

1. Bardhan, P. K., 1965, 'Equilibrium growth in the international economy', *Quarterly Journal of Economics*, **79**, 455-64.
2. Bertrand, Trent J., 1971, 'The gains from trade an analysis of steady-state solutions in the open economy', mimeographed, Johns Hopkins University
3. Drandakis, Emanuel M, 1963, 'Factor substitution in the two-sector growth model', *Review of Economic Studies*, **30**, 217-28
4. Johnson, Harry G., 1971, 'Trade and growth a geometrical exposition', *Journal of International Economics*, **1**, 83-101
5. Oniki, H, and Uzawa, H, 1965, 'Patterns of trade and investment in a dynamic model of international trade', *Review of Economic Studies*, **32**, 15-38
6. Stiglitz, Joseph E, 1970, 'Factor price equalization in a dynamic economy', *Journal of Political Economy*, **78**, 456-88.
7. Uzawa, Hirofumi, 1963, 'On a two-sector model of economic growth II', *Review of Economic Studies*, **30**, 105-18.
8. Vanek, Jaroslav, 1971, 'Economic growth and international trade in pure theory', *Quarterly Journal of Economics*, **85**, 377-90.

A MODEL OF THE INFLATION CYCLE IN A SMALL OPEN ECONOMY[1]

By B. L. SCARFE

1. Introduction

It is a fairly well-known proposition that a small open economy has ve
little freedom in its choice of an appropriate point on its long-run 'inflatio
unemployment' trade-off curve under fixed exchange rates. Indeed, in th
extreme case, the equilibrium rate of money wage inflation *and* the equil
brium rate of unemployment in a small open economy with a trade-o
curve are entirely determined by the rate of world price inflation unde
fixed exchange rates.[2] While this result extends without difficulty to th
flexible exchange rate case as long as the rate of appreciation or deprecia
tion is held constant, it is by varying its rate of appreciation or depreciatio
by appropriate monetary, fiscal, and commercial policies that a small ope
economy can partially insulate itself from fluctuations in the rate of worl
price inflation.

It is the purpose of this paper to explore the dynamic behaviour of a
simple economy for which this central proposition holds. It will be show
that such an economy is likely to experience an *inflation cycle* in which the
boom phase is followed by a phase of inflationary recession, then by a
slump phase and, finally, by a phase of dis-inflationary expansion.

For any small open economy on a fixed exchange rate there are four basic mechanisms by which foreign inflation tends to become domestic inflation.[3] These four inflation transmission mechanisms are (*a*) cost-push effects, (*b*) demand-pull effects, (*c*) institutional or non-market effects, and (*d*) monetary effects. Cost-push effects occur directly through the domestic

[1] While the author's ideas on the topic of inflation took shape while he was a member of the research division of the Prices and Incomes Commission in Ottawa, this paper should in no way be construed to represent the views of the Commission By the same token, the final report of the Prices and Incomes Commission, *Inflation, Unemployment and Incomes Policy*, Ottawa, Information Canada, 1972, should not be construed to represent the views of the current author Indeed, as this paper makes clear, those views are entirely misrepresented in footnote 7, p. 61 of the Report when read in context with the text above it. The author is indebted to A M C Waterman, who first stimulated his interest in models of the type described herein, and to N E. Cameron, J G Cragg, J S Flemming, and D Hum, for valuable comments and advice The basic model underlying this paper bears an essential similarity to the model outlined in Waterman, 'A simple Keynesian alternative to Neher's model', in N Swan and D. A Wilton, *Inflation and the Canadian Experience*, Queen's University, Kingston, 1971

[2] See, for example, T. W. Swan, 'Economic control in a dependent economy', *Economic Record*, vol 36, 1960, pp 51–66, and A. M. C. Waterman, 'Some footnotes to the "Swan Diagram" or "How dependent is a dependent economy?"', ibid , vol. 42, 1966, pp. 447–64

[3] For expansion on these four mechanisms, see B. L Scarfe, *Price Determination and the Process of Inflation in Canada*, Ottawa, Information Canada, 1972.

pricing equations when the prices of imported commodities that enter either further production or the domestic consumption bundle rise Demand-pull effects occur whenever foreign inflation leads to an increase in the demand for domestic exports or import-competing products by making their foreign substitutes temporarily more expensive This expansion in demand increases the price of domestically produced commodities, thereby tending to stabilize the relative price of foreign and domestically produced commodities. In addition, the expansion of demand for domestic output reduces unemployment which increases the rate of money wage inflation, with ultimate effects on domestic price inflation. Institutional linkages transmit inflation via direct wage emulation within the international union structure and via direct price setting within the structures of multi-national corporations Finally, monetary effects result from the fact that with elastic capital flows it is difficult, if not impossible, to continue sterilizing the inflows of foreign exchange that are generated by the improvement in the trade balance inherent in the demand-pull effects Ultimately, therefore, these inflows may well induce an increase in the domestic money supply.

2. The formal structure of the model

The basic model outlined herein cuts through these several transmission mechanisms by ignoring the commodity market and the money market and concentrating on the labour market The central endogenous variables of the model are the actual domestic unemployment rate (u), the potential domestic unemployment rate (u_e), the actual rate of domestic wage inflation (w), and the expected rate of domestic wage inflation (w_e) The potential domestic unemployment rate will be seen later to be the 'target' rate towards which the actual rate is moving at any point of time Let it be assumed that the potential domestic unemployment rate is an increasing function of the economy's degree of 'non-competitiveness' This non-competitiveness may be measured by (the natural logarithm of) the economy's 'cost ratio', where the cost ratio is defined to be the ratio of normal unit labour costs (the wage rate divided by the normal level of labour productivity) to foreign prices adjusted for the exchange rate (the level of foreign prices divided by the price of domestic currency) More specifically, let it be assumed that the potential domestic unemployment rate increases whenever normal unit labour costs grow more quickly than adjusted foreign prices (that is, whenever the cost ratio is growing), and that the potential domestic unemployment rate decreases whenever normal unit labour costs grow more slowly than adjusted foreign prices (that is, whenever the cost ratio is falling) Since the growth rate of normal unit labour costs is equal to the rate of domestic wage inflation (w) less the rate

of increase (ξ) of normal labour productivity, and since the growth rate of foreign prices when expressed in terms of domestic currency is equal to the rate of foreign price inflation (P_f) less the rate of appreciation (π) of the price of domestic currency, the expression $w-\xi-P_f+\pi$ may be taken to be the growth rate of the cost ratio

Let c be the natural logarithm of the cost ratio. Then the relationship between the potential unemployment rate and the cost ratio may be written as

$$u_e = \Omega(c) \quad \text{and} \quad Du_e = \Omega'_c\ Dc, \tag{1}$$

where $\Omega(c)$ is a continuous differentiable function with $\Omega'_c(c) > 0$, and where D is the operator d/dt But since Dc is the growth rate of the cost ratio, one has

$$Dc = w-\xi-P_f+\pi \tag{2}$$

Combining (1) and (2) yields the relationship

$$Du_e = \Omega'_c.(w-\xi-P_f+\pi) \tag{3}$$

In economic terms, these expressions assert that the cost ratio and the potential unemployment rate both increase (decrease) whenever normal unit labour costs grow more quickly (slowly) than foreign prices when expressed in terms of domestic currency In addition, since the *sign* of $w-\xi-P_f+\pi$ is always opposite to the *sign* of the growth rate of the profit margin (if the profit margin is itself assumed to be positive) these expressions also assert that the potential unemployment rate increases whenever the profit margin contracts and decreases whenever the profit margin expands

The potential unemployment rate is, however, not the same as the actual unemployment rate Nevertheless, it may be assumed that the actual unemployment rate (u) adjusts to the potential unemployment rate (u_e) in a lagged fashion If this lag is taken to be of a simple exponential form with speed of response $\alpha > 0$, then one may write

$$(D+\alpha)u = \alpha u_e \tag{4}$$

In addition to expressions (3) and (4), the model also incorporates an 'inflation-unemployment' trade-off function or Phillips curve This function relates the rate of domestic wage inflation (w) to the unemployment rate (u) and to the expected rate of domestic wage inflation (w_e) Let $\phi(u)$ be a continuous twice-differentiable short-run trade-off function with $\phi'_u(u) < 0$, $\phi''_u(u) > 0$, and let $0 \leqslant b \leqslant 1$ be the expectations coefficient Then w may be assumed to adjust to $\phi(u)+bw_e$ with some institutional lag If this lag is taken to be of a simple exponential form with speed of response $\mu > 0$, then one may write the complete trade-off function as

$$(D+\mu)w = \mu[\phi(u)+bw_e] \tag{5}$$

Finally, to complete the model, it may be assumed that the expected rate of wage inflation adapts to the actual rate of wage inflation with some

expectational lag. If this lag is taken to be of a simple exponential form with speed of response $\rho > 0$, then one may write this 'adaptive expectations' hypothesis as

$$(D+\rho)w_e = \rho w \tag{6}$$

3. Some basic properties of the model

The model outlined consists of four first-order differential equations, expressions (3) to (6) inclusive. These equations contain two non-linear elements in the form of the functions Ω_c' and $\phi(u)$. However, given the derivative conditions placed upon $\phi(u)$, there will exist a unique stationary point to the system for any exogenously given configuration of P_f, π, and ξ.[1] This stationary point is the solution

$$w^* = P_f - \pi + \xi, \quad \phi(u^*) = (1-b)w^*, \quad w_e^* = w^*, \quad \text{and} \quad u_e^* = u^*, \tag{7}$$

where P_f, π, and ξ are assumed to be given exogenously.[2] Notice that w^*, u^*, w_e^*, and u_e^* are all uniquely determined from these three exogenous variables. Moreover, since the rate of increase (ξ) in normal labour productivity is largely determined by long-run technological developments, it is clear that the rate of foreign price inflation (P_f) is the fundamental exogenous variable in the model if the rate of appreciation (π) of the price of domestic currency is held constant. This is particularly the case under a regime of fixed exchange rates, where $\pi = 0$. It is therefore clear that the only way in which foreign inflationary pressures can be offset in the long-run is by appreciating the domestic currency in a one-for-one manner with changes in the world price level. Since this is impossible if the exchange rate remains fixed, a flexible exchange rate emerges as the primary instrument for combating imported inflationary pressures in a small open economy.

The dynamic behaviour of the economy when it is not at the stationary point can be explored by linearizing the model around this point. Let

[1] It should be noted that although Ω_c' is evaluated for a particular value of c, where c depends upon an integral of $w - \xi - P_f - \pi$, it does *not* follow that the value of c at the stationary point is not uniquely determined. While the value of c depends upon the whole past history of u, ξ, P_f, and π, *if a stationary point is attained* the value of u must have adjusted in response to u, and hence to ξ, P_f, and π in such a way as to make c unique. Of course, alternative configurations of $P_f - \pi - \xi$ must lead to alternative stationary points with different values of w, u and c. It is, however, worth noting that the author is aware that he is at this point skirting around some difficult mathematical problems associated with the integral adjustment process underlying the definition of c.

[2] If $b = 1$, the equilibrium condition $(1-b)u^* = \phi(u^*)$ implies that the solution for u^* must be replaced by $\phi(u^*) = 0$. In this case, u^* is said to be 'the natural rate of unemployment'. The existence of a natural rate of unemployment implies that there is no long run trade off relationship between inflation and unemployment. In this case it is not the rate of foreign price inflation which constrains the economy's choice of an appropriate rate of unemployment in the long-run, but rather it is the internal characteristics of the labour market which provide the constraint. Thus the unemployment part of the central proposition enunciated at the beginning of this paper is only non trivial in the case where the expectations coefficient, b, is less than unity. Since there are no very persuasive reasons for supposing that b is equal to unity, and since the resulting dynamic behaviour is hardly affected if b is unity in any case, it will be assumed henceforth in this paper that $b < 1$.

$u_e = u_e^* + \bar{u}_e e^{\gamma t}$, $u = u^* + \bar{u}e^{\gamma t}$, $w = w^* + \bar{w}e^{\gamma t}$, and $w_e = w_e^* + \bar{w}_e e^{\gamma t}$ In addition, let $\phi'_* < 0$ be the value of $\phi'(u)$ when $u = u^*$, and let $\Omega'_* > 0$ be the value of Ω'_e when the system is at the stationary point Then the linearization procedure yields the matrix equation [1]

$$\begin{bmatrix} \gamma & 0 & -\Omega'_* & 0 \\ -\alpha & \gamma+\alpha & 0 & 0 \\ 0 & -\mu\phi'_* & \gamma+\mu & -\mu b \\ 0 & 0 & -\rho & \gamma+\rho \end{bmatrix} \begin{bmatrix} \bar{u}_e \\ \bar{u} \\ \bar{w} \\ \bar{w}_e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{8}$$

Setting the determinant of the 4×4 matrix equal to zero, one finds the characteristic equation to be the quartic

$$\gamma^4+(\alpha+\mu+\rho)\gamma^3+[\alpha(\mu+\rho)+\mu\rho(1-b)]\gamma^2+ \\ +[\alpha\mu\{\rho(1-b)-\Omega'_*\phi'_*\}]\gamma - \alpha\mu\rho\Omega'_*\phi'_* = 0 \tag{9}$$

Applying Descartes' Rule of Signs, it is easy to see that this polynomial cannot possess any real non-negative roots since all of its coefficients are positive Thus, if the system is unstable in the neighbourhood of the stationary point, it must be cyclically unstable However, applying the Routh–Hurwitz conditions, it may be shown that all the characteristic roots are either real negative numbers or complex numbers with negative real parts if and only if [2]

$$[\alpha+\mu+\rho b+\Omega'_*\phi'_*][(\alpha+\rho)\rho+\mu\rho b-\alpha\mu\Omega'_*\phi'_*] > \rho b(\alpha+\mu+\rho)^2 \tag{10}$$

It follows that if this condition holds the system is stable in the neighbourhood of the stationary point (that is, it is locally stable) Since the second term on the left-hand side and the term on the right-hand side of expression (10) are necessarily positive, it follows that it is necessary for local stability that $\alpha+\mu+\rho b+\Omega'_*\phi'_* > 0$ The adjustment parameters α, μ, and ρb must, in combination, be large enough to ensure that the positive feedback from w to u in Ω'_* combined with the negative feedback from u to w in ϕ'_* does

[1] Since (for example) $Du = D(u^*+\bar{u}e^{\gamma t}) = \gamma\bar{u}e^{\gamma t}$, and since the linearization procedure implies the approximation $\phi(u^*+\bar{u}e^{\gamma t}) \approx \phi(u^*)+\phi'_*(u^*+\bar{u}e^{\gamma t}-u^*)$, one may write

(a) $\gamma\bar{u}_e = \Omega'_*\bar{w}$, (b) $(\gamma+\alpha)\bar{u} = \alpha\bar{u}_e$,

(c) $(\gamma+\mu)\bar{w} = \mu(\phi'_*\bar{u}+b\bar{w}_e)$, and (d) $(\gamma+\rho)\bar{w}_e = \rho\bar{w}$

from expressions (3) to (6), respectively The four equations (a) to (d) correspond to the matrix equation given in the text above

[2] The Routh–Hurwitz conditions state that the characteristic roots of the quartic equation $k_4\gamma^4+k_3\gamma^3+k_2\gamma^2+k_1\gamma+k_0 = 0$ all have negative real parts if and only if the four principal minors of

$$\begin{bmatrix} k_1 & k_0 & 0 & 0 \\ k_3 & k_2 & k_1 & k_0 \\ 0 & k_4 & k_3 & k_2 \\ 0 & 0 & 0 & k_4 \end{bmatrix}$$

are all positive, provided (as may legitimately be assumed) $k_0 > 0$ (See, for example, J V Uspensky, *Theory of Equations*, New York, McGraw-Hill, 1948, p 304) Given the sign constraints imposed on the parameters of the model discussed herein it can be shown that these conditions reduce to the single inequality given in the text above

not lead to an explosive oscillation around the stationary point. However, these coefficients, and particularly the expectations combination, ρb, must not be so large as to reverse the sign of the inequality given in expression (10). For if this occurs, it will also generate explosive oscillations around the stationary point. Thus, there are two possible reasons why instability in the form of explosive oscillations may occur, though the oscillatory feature itself is a consequence of the positive and negative feedback combination mentioned above.

4. The inflation cycle in a simplified case

On the assumption that condition (10) holds, the system will be locally stable. However, from any point in the neighbourhood of the stationary solution, this solution may well be approached cyclically. This may be illustrated by means of a phase diagram (Fig. 1). The phase diagram has been constructed on the simplifying assumptions that the expectational lag and the unemployment lag are both very short relative to the institutional lag in the wage adjustment process, so short indeed that they can safely be ignored. In this case, $w_e = w$ and $u_e = u$ always and the equation system reduces to

$$(D+\mu)w = \mu[\phi(u)+bw] \quad \text{and} \quad Du = \Omega_c'[w-\xi-P_f+\pi] \tag{11}$$

In the phase diagram, the actual domestic unemployment rate (u) is measured on the horizontal axis while the actual rate of domestic wage inflation (w) is measured on the vertical axis. The downward-sloping singular curve labelled $Dw = 0$ represents the locus of all points at which there is no change in the rate of wage inflation. From expression (11) it is evident that if $\phi(u) > (1-b)w$, then $Dw > 0$. This is represented in the phase diagram by the arrows pointing vertically upwards in the quadrants labelled (*a*) and (*b*). On the other hand, if $\phi(u) < (1-b)w$, then $Dw < 0$, which is represented in the phase diagram by the arrows pointing vertically downwards in the quadrants labelled (*c*) and (*d*). The horizontal singular curve labelled $Du = 0$ represents the locus of all points at which there is no change in the unemployment rate. From expression (11) it is evident that if $w > P_f-\pi+\xi$, then $Du > 0$. This is represented in the phase diagram by the arrows pointing horizontally to the right in the quadrants labelled (*b*) and (*c*). On the other hand, if $w < P_f-\pi+\xi$, then $Du < 0$, which is represented in the phase diagram by the arrows pointing horizontally to the left in the quadrants labelled (*d*) and (*a*).

In any particular quadrant of the phase diagram, the *combination* of arrows explains the direction of movement in the unemployment rate and the rate of wage inflation in that quadrant. When these movements are combined into a single story it becomes evident that oscillations in the unemployment rate and the rate of wage inflation are likely. These

If this cyclical momentum is to be avoided in such an economy, it is essential that the rate of appreciation of the domestic currency be allowed to increase when world price inflation accelerates and to decrease when world price inflation decelerates Indeed, the only effective tool for insulating such an

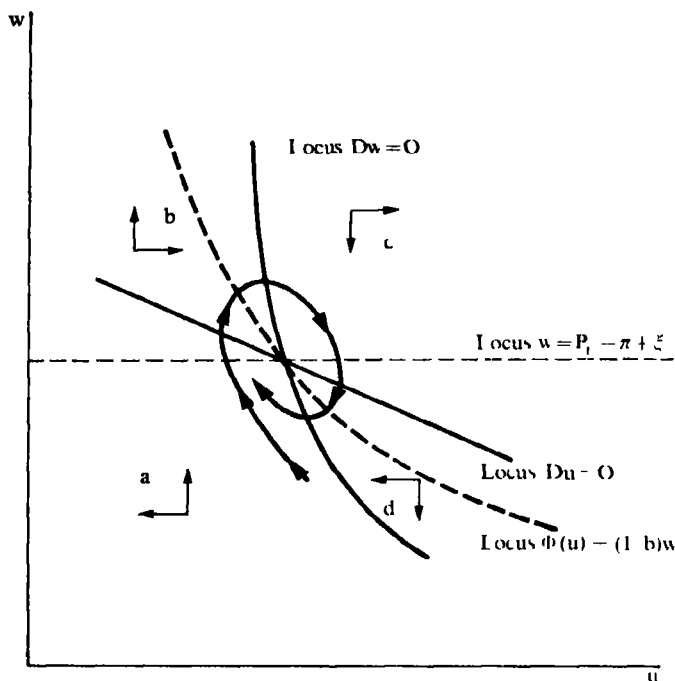



FIG 2 Phase diagram for general case

economy from fluctuations in the rate of world price inflation is an appropriately managed flexible exchange rate. Without this tool, it would be exceedingly difficult, if not impossible, to maintain either the rate of wage inflation or the rate of unemployment close to their desired or target levels

The management of the exchange rate is, of course, interdependent with monetary and fiscal policy The point made in the previous paragraph does not imply that these policies are powerless. On the contrary, not only is their co-ordinated power enhanced under flexible exchange rates, but also they are the primary instruments through which the exchange rate may be appropriately managed The point is that to maintain a fixed exchange rate (or even a fixed rate of appreciation or depreciation) in the face of fluctuations in the rate of world price inflation is to make it exceedingly difficult to stabilize the economy For, on the one hand, monetary policy ceases to be an internal stabilizing device when it is tied to the maintenance of a rigid exchange rate, while, on the other hand, fiscal policy may be

constrained by the unwillingness of the authorities to withstand large-scale changes in the public debt

While it is possible (when capital flows are elastic) to co-ordinate monetary and fiscal policy under fixed exchange rates in such a way as to maintain for some period of time a different combination of u and w than world conditions would dictate, to do so is to try to stabilize the economy with the big lever tied behind one's back. For example, continuous monetary tightening and fiscal easing could maintain u at a level low enough for $w - \xi > P_f$, while continuous fiscal tightening and monetary easing could maintain u at a level high enough for $w - \xi < P_f$. However, the eventual outcome of both of these policy combinations will be an abrupt change in the exchange rate when external pressures force (in the former case) a depreciation or (in the latter case) an appreciation of the domestic currency.[1]

Even if it turns out to be impractical to suggest either continuous appreciation or continuous depreciation through time,[2] so that an economy must eventually accept the rate of world price inflation in the long-run, it is still the case that an exchange rate that is allowed to fluctuate (perhaps widely) around some long-run norm is the strongest lever available for offsetting fluctuations in the rate of world price inflation and preventing them from causing fluctuations in the domestic unemployment rate. In this manner, it is possible to reduce considerably the amplitude of cyclical fluctuations in u and perhaps also to reduce the *average* rate of unemployment over time.[3]

University of Manitoba, Winnipeg

[1] This latter outcome is well illustrated by the events leading up to the appreciation of the Canadian dollar in 1970. The former outcome might well occur if the Canadian dollar were pegged at its current high level and the United States wage and price control programme turns out to be reasonably successful. For in such a situation one must either allow greater unemployment in Canada or eventually depreciate the currency.

[2] There is some evidence that these options are considered to be impractical by Canadian financial authorities. It may well be that this is a consequence of the (probably incorrect) belief that institutional or non market linkages are the most important inflation transmission mechanism from the United States to Canada, linkages which can hardly be offset by exchange rate variations in any case. There is also some evidence that these same authorities attempt in a somewhat mercantilistic (and self-defeating) way to keep the Canadian economy in a phase of the inflation cycle where profit margins are expanding, the cost ratio is falling, and the unemployment rate is falling. Of course, to remain in any one phase perpetually is impossible, no matter how undesirable other phases (such as inflationary recession) might appear to be.

[3] If the Phillips curve relationship is non-linear with $\phi''_u(u) > 0$, the impact on w of employment lost in a cyclical downswing does not offset the impact on w of employment gained in a cyclical upswing in a one-for-one manner. Thus the greater the amplitude of cyclical fluctuations in u the higher must the *average* u be if w is to remain constant on average. Moreover, the *inefficiency* involved in letting u follow a 'stop-go' cycle rather than a more stable course may well be enhanced if the productive quality of the employment gained in a boom phase is lower than the productive quality of the employment lost in a slump phase. For one then has to give up a larger quantity of more productive employment to get a smaller quantity of less productive employment in order to keep u constant on average over cyclical swings in u.

SALES VERSUS INCOME TAXES: A PEDAGOGIC NOTE

By BRIAN MOTLEY[1]

DISCUSSIONS of the impact of fiscal policy on aggregate demand in most macroeconomic textbooks typically assume that all taxes are levied on the *income* of households. However, many governments not only derive a large proportion of their total revenues from *sales* taxes but also vary the levels of these taxes as an instrument of stabilization policy. No harm would be done by this textbook simplification if income and sales taxes yielding equivalent revenue had identical effects on aggregate demand. This note seeks to demonstrate that in an important class of cases (that in which the change is expected to be a temporary one) this will not be the case

In addition, few texts draw attention to the implications for fiscal policy of the 'new theories of the consumption function' even though the latter are discussed in some detail in the chapters devoted to the determinants of consumer expenditures[2] Both the permanent income and life-cycle hypotheses, for example, imply that an income-tax increase will cause households to reduce their current consumption only to the extent that it leads them to scale down their estimates of total lifetime wealth Hence, if a tax increase is widely assumed to be temporary, these hypotheses lead to the conclusion that its effects on aggregate demand are likely to be minimal because its impact on total wealth will be relatively much smaller than that on current income[3] This argument is well known in the journal literature[4] but apparently has not yet trickled down to the major undergraduate texts

[1] The author wishes to thank Stuart Burness, Hirofumi Shibata, and Sherwin Rosen for helpful comments on an earlier draft of this paper Naturally, he accepts all responsibility for the finished product

[2] Edward Shapiro's widely used text, *Macroeconomic Analysis*, is a good example Although Shapiro devotes considerable space to the permanent income hypothesis (pp 159–63, 178) in his chapter on consumption, his theoretical analysis of fiscal policy (Chapter 14) maintains the assumption that household spending depends only on current disposable income

[3] This argument suggests that governments (such as that of Great Britain, for example) which pursue a 'Keynesian' fiscal policy of changing income-tax rates very frequently, are likely to find that they soon run into diminishing returns In Friedman's terms, year-to-year variations in tax liabilities come to be regarded by the public as largely 'transitory' changes which do not affect their estimates of lifetime wealth and as a result have little or no influence on spending. For similar reasons, the argument casts doubt on the usefulness of a policy of giving an American President stand-by authority to make temporary changes in income-tax rates

[4] See, for example, Robert Eisner, 'Fiscal and monetary policy reconsidered', *American Economic Review*, Dec 1969, 59, pp. 897–905. Incidentally, this paper contains the only reference I have been able to uncover to the effect analysed in this paper

However, the argument that a temporary increase in income taxes will have little effect on current consumption does not carry over from income taxes to other classes of levies. As I shall show in this note a temporary increase in *sales* taxes is likely to have a greater effect on current consumption than an income-tax increase yielding the same revenue In fact, I shall show that it is possible to devise a combined policy which raises sales taxes, reduces income taxes and is contractionary even though government revenue remains unchanged [1]

Although sales and income taxes fall with differential impact on different households, all distribution effects will be ignored Hence, the analysis will be conducted in terms of an 'average household' This household chooses a consumption programme to maximize lifetime utility subject to an over-all wealth constraint I shall assume that this multiperiod decision problem may be collapsed[2] into a two-period problem involving only 'the present' and 'the future'. However, the present does not necessarily correspond to the current income period Instead, the two periods are defined such that a 'temporary' tax change affects only the present and not the future. I assume throughout that both present and future consumption are normal (that is, non-inferior) goods [3] In addition, all multiplier effects are ignored, thus the analysis is concerned only with the 'first round' impacts of policy changes

Fig 1 depicts the initial position of an average household *OP* and *OF* represent present and future (after-income-tax) income respectively The household is able to borrow or lend at a real rate of interest given[4] by the slope of *XZ*. Hence, *OXZ* represents the set of feasible consumption programmes open to the household. Given this constraint, the household chooses the programme represented by the point *A*, consuming *OH* in the present and *OG* in the future.

Assume now that the government imposes a tax on present consumption, that is, it levies a sales tax on consumer goods which is expected to

[1] Following publication of Kaldor's *An Expenditure Tax*, a debate developed on the relative effects of an income tax and an expenditure tax on saving This literature is summarized in Richard Goode, *The Individual Income Tax* (Brookings Institution, 1964), Chapter 3 This literature, however, was concerned exclusively with the long-run effects of a *permanent* change in the method of raising tax revenue Also, much of it pre-dated the permanent-income hypothesis and, hence, did not incorporate the more modern approach to the determination of consumption and saving In this note I treat the case of a *temporary* tax change and cast my analysis within the framework of the new theories of the consumption function

[2] For discussions of the conditions under which this aggregation is possible see the contributions of Green, Hicks, and Blackorby, *et al*

[3] Although obviously plausible, this assumption is not essential to the argument and is made only for stylistic reasons

[4] If r is the relevant real rate of interest, the ratio OX/OZ is equal to $1+r$ However, since both present and future consumption are 'composite goods', r will not generally correspond with any observable market rate of interest

be temporary. Such a policy increases the price of present consumption relative to that of future consumption and, hence, worsens the terms on which additions to present consumption may be 'bought' by reducing future consumption.[1] Before the sales tax was introduced, a one-dollar

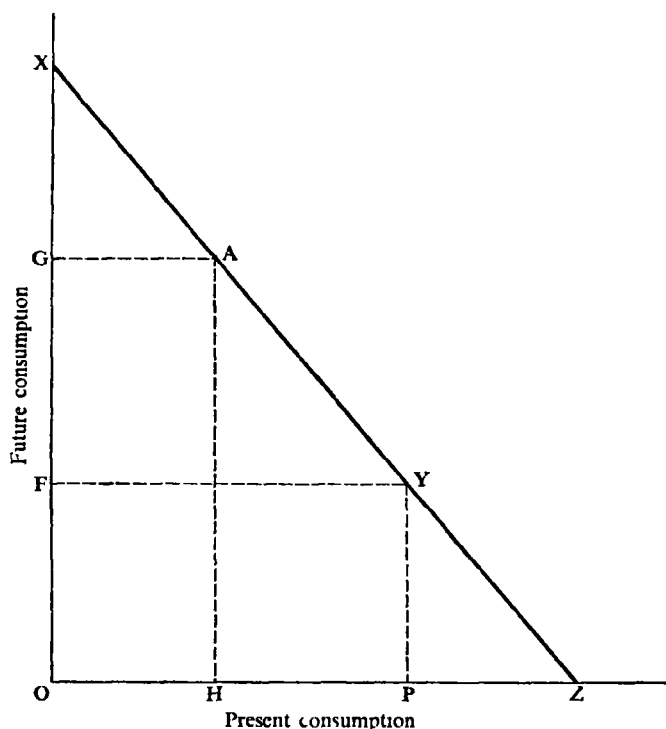



FIG 1

increase in present consumption necessitated a reduction in future consumption of $1+r$ dollars. The imposition of a sales tax on present consumption at a rate of $100t$ cents per dollar increases the required future sacrifice to $(1+r)(1+t)$ dollars This may be seen by examining the wealth-constraint of the household before and after the imposition of the sales tax.

Before the imposition of the sales tax the household's wealth constraint may be written (in obvious notation)

$$Y_P+Y_F/(1+r) = C_P+C_F/(1+r)$$

The quantity on the left side of this equation is the household's present

[1] A permanent sales tax will not have this 'relative price effect' In this case, the effects of a sales tax are precisely analogous to those of an income tax although sales and income taxes having the same effects on the household's budget constraint may not yield the same current revenue to the government

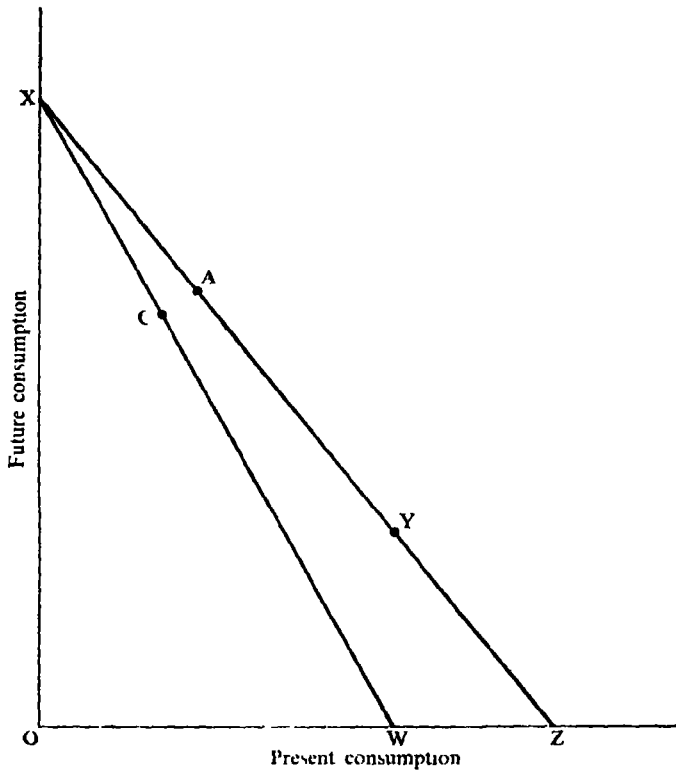



FIG 2

wealth. The terms on which future consumption can be increased at the expense of present consumption are given by

$$\frac{\partial C_F}{\partial C_P} = -(1+r)$$

After the sales tax is imposed the wealth constraint becomes[1]

$$Y_P + Y_F/(1+r) = (1+t)C_P + C_F/(1+r)$$

and hence
$$\frac{\partial C_F}{\partial C_P} = -(1+r)(1+t).$$

A given sacrifice of present consumption now makes possible a larger increase in future consumption

The effect of this tax change on the householder's situation is illustrated in Fig. 2 As in Fig 1 OXZ represents the pre-tax budget constraint The post-tax constraint must pass through the point X and have slope

[1] I am ignoring the question of what the government does with its additional revenues If it uses them to make additional transfer payments (in the form either of cash or of extra government services), the average household's present income will be increased The case in which the revenues are used to effect an equal reduction in income tax is treated below.

$-(1+r)(1+t)$ OXW represents this constraint. The distance OW represents the quantity $C_P+C_F/(1+r)(1+t)$ and hence the tax rate, t, is given by the ratio WZ/OW Given this new constraint the household chooses a consumption plan represented by the point C.

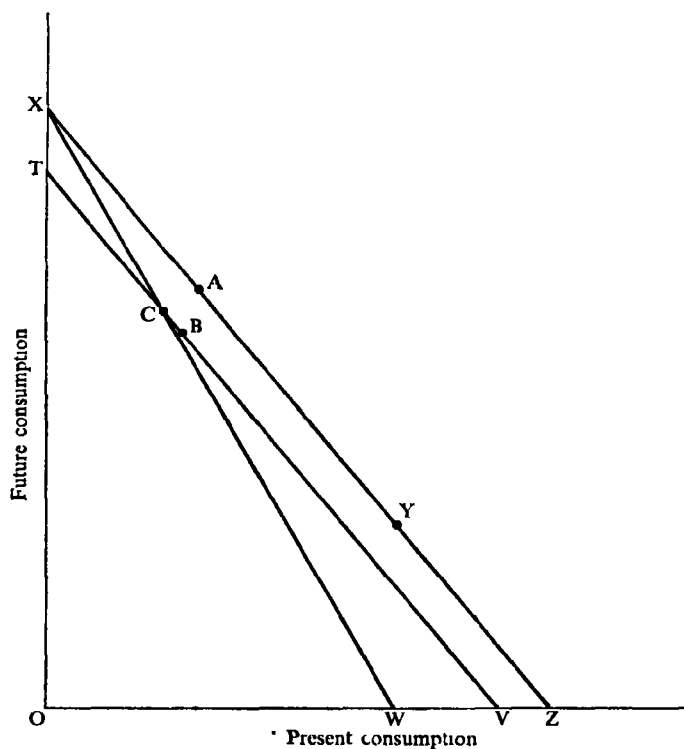



FIG 3

I now wish to show that a temporary sales tax leads to a reduction in current consumption which is larger than that produced by a temporary income tax yielding the same revenue [1]

Consider Fig 3 The line TV is constructed to pass through the point C and to be parallel to XZ Then the distance VZ represents the receipts derived from the sales tax To prove this note that[2]

$$OZ = Y_P+Y_F/(1+r) = C_P^*(1+t)+C_F^*/(1+r)$$

[1] In the (unlikely) event that present consumption is an inferior good, an income tax will be expansionary. In this case the following argument may be used to show that the sales tax is less expansionary than the income tax

[2] The starred variables denote the values of C_P and C_F which maximize lifetime utility when the sales tax is in effect

and that, since TV passes through the point C and has slope $-(1+r)$,

$$OV = C_P^*+C_F^*/(1+r).$$

Hence, $VZ = OZ-OV = tC_P^*$ which is the yield of a sales tax at rate t levied on present consumption of C_P^*

Thus, a temporary income tax which yields the same revenue will be one which reduces present income by an amount VZ The imposition of such a tax will result in the household's being faced with the wealth constraint OTV. In such a situation it will choose the consumption programme represented by the point B As long as present consumption is a normal good, this programme will involve a lower level of present consumption than does the programme A

However, present consumption is necessarily lower at C than at B One simple proof of this fact is as follows. Under the income tax, the household will choose a consumption programme represented by a point on TV. We know, however, that C is preferred to all the other programmes on TC since it was chosen when those other programmes were available [1] As a result, we know that (under the income tax) the household will choose a programme on CV Hence, the chosen programme, B, necessarily involves a higher[2] level of current consumption than does C This argument, together with the assumption of noninferiority, establishes both that sales taxes are contractionary (C lies to the left of A) and that their impact is greater than that of income taxes yielding equal revenue (C lies to the left of B).

The change in current consumption resulting from the imposition of a temporary sales tax may be divided into a *wealth effect* and a *substitution effect*. The wealth effect is represented by the shift from A to B and is negative so long as current consumption is not inferior. The substitution effect,[3] which takes the household from B to C, is also negative [4] By contrast, an income tax yielding the same revenue produces a wealth effect only and, hence, is less contractionary than a sales tax.

Finally, I wish to demonstrate that a suitable combination of a sales tax increase and an income-tax reduction may be contractionary even though it leaves the government's revenue unchanged. To establish this proposition consider Fig. 4.

As before, OXZ represents the pre-sales-tax wealth constraint and OXW the constraint after the tax is imposed. Similarly A and C represent the

[1] Under the sales tax the budget constraint was OXW and programmes on TC were feasible under this constraint. However, the programme chosen was that represented by C

[2] We ignore the theoretical possibility that B and C coincide.

[3] To be precise, this is a substitution effect in the sense of Slutsky. For a discussion of the distinction between Slutsky and Hicksian substitution effects see Bilas (pp. 72–8).

[4] The change in *future* consumption consists of a negative wealth effect and a *positive* substitution effect. Hence, the net effect is ambiguous.

consumption programmes chosen before and after the sales tax is introduced Now suppose that after imposing the sales tax, the government reduces *income* taxes by an amount such that the originally chosen programme (A) is still open to the household. This requires an increase in present (after-tax) income which is sufficient[1] to shift the wealth constraint outward from OXW to ORS.

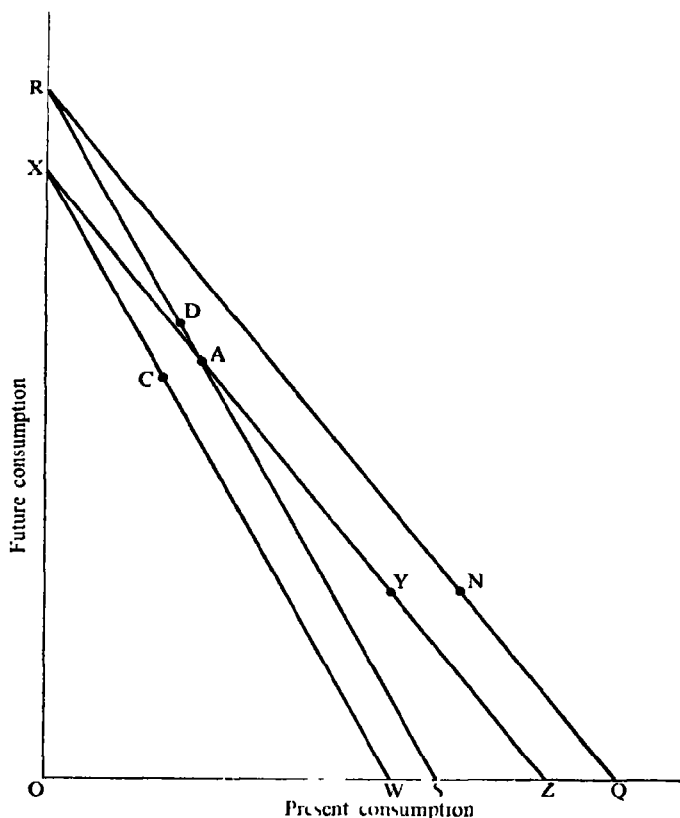



FIG. 4

If the household actually were to choose the consumption programme represented by A, the revenue from the sales tax would exactly offset the reduction in income-tax receipts so that the government's total revenue would be unaffected by this policy In fact, however, the household will *not* choose A but will prefer D which lies to the left[2] of A (that is, involves

[1] By constructing RQ parallel to XZ we find that the necessary reduction in income taxes amounts to ZQ, that is, the original income point must shift from Y to N The proof is left to the reader.

[2] In the original (no-tax) situation, programme A was chosen when points on AS were available Hence, when the budget constraint is ORS, the chosen programme necessarily lies on RA.

a lower level of current consumption than A). As in the previous case, the shift from A to D represents a Slutsky substitution effect, which is always negative. Thus, we conclude that this policy, which shifts the budget constraint from OXZ to ORS, is contractionary

On the other hand, when the budget constraint is ORS and programme D is chosen, government revenue is less[1] than in the original situation Thus, we have found a government policy which *reduces* government revenue and yet is *contractionary*. Hence, if government revenue is to remain unchanged, the income-tax remission must be *less* than ZQ A combination of an income-tax remission somewhat less than ZQ together with the temporary sales tax will be even more contractionary than the policy depicted in Fig 4 and will leave government revenue unchanged [2]

Conclusion

The argument of this brief paper suggests that income taxes and sales taxes may each perform a different function in fiscal policy Short-term stabilization policy is best performed by the use of sales-tax changes. Since these changes are most potent when they are expected to be temporary, the government might consider it desirable to announce a date for their removal when they are introduced. Long-term shifts in the proportion of the nation's wealth which is left in the hands of private spenders, on the other hand, are probably best made by changes in the rate of personal income tax.

University of Kentucky, Lexington

REFERENCES

1. BILAS, RICHARD A., *Macroeconomic Theory*, second edition, McGraw-Hill (1971).
2. BLACKORBY, C, LADY, G, NISSEN, D, and RUSSELL, R R, 'Homothetic separability and consumer budgeting', *Econometrica* (May 1970) **38** (3).
3. FRIEDMAN, MILTON, *A Theory of the Consumption Function*, Princeton University Press (1957)
4. GOODE, RICHARD, *The Individual Income Tax*, Brookings Institution (1964)
5. GREEN, H A JOHN, *Aggregation in Economic Analysis*, Princeton University Press (1964).
6. HICKS, J R., *Value and Capital*, second edition Oxford University Press (1946).
7. KALDOR, N, *An Expenditure Tax*. Allen and Unwin (1955)
8. SHAPIRO, EDWARD, *Macroeconomic Analysis* second edition, Harcourt, Brace and World (1970).
9. SLUTSKY, EUGEN A, 'On the theory of the budget of the consumer', reprinted in Stigler, G. J and Boulding, K. E, *Readings in Price Theory*, Richard D Irwin (1952)

[1] Because current consumption, and hence the yield from the sales tax, is lower than when programme A is chosen

[2] The reader who is unconvinced by this verbal and geometric argument will find a mathematical proof in the appendix to this note.

APPENDIX

In this appendix I prove that a combination of a sales-tax increase and an income-tax decrease which leaves government revenue unchanged will reduce present consumption

The consumer's decision problem is as follows

Maximize $U(C_P, C_F)$.

Subject to $Y_P - T + Y_F/(1+r) = C_P(1+t) + C_F/(1+r)$,

where T represents income taxes levied on present income and all other notation is the same as in the text above.

The first-order conditions for a solution require

$$U_1 = \lambda(1+t),$$
$$U_2 = \lambda/(1+r).$$

Totally differentiating these conditions and the budget constraint while holding Y_P, Y_F, and r constant yields the following set of linear equations:

$$u_{11}\,dC_P + u_{12}\,dC_F - (1+t)\,d\lambda = \lambda\,dt, \quad (1)$$
$$u_{12}\,dC_P + u_{22}\,dC_F - [1/(1+r)]\,d\lambda = 0, \quad (2)$$
$$-(1+t)\,dC_P - [1/(1+r)]\,dC_F = C_P\,dt + dT. \quad (3)$$

We wish to consider changes in t and T which leave government revenues unchanged. This implies the additional constraint

$$d(tC_P) + dT = 0$$

which may also be written

$$C_P\,dt + dT = -t\,dC_P. \quad (4)$$

Substitution of (4) into (3) gives

$$dC_P + 1/(1+r)\,dC_F = 0. \quad (3a)$$

With no loss of generality we may assume[1] $t = 0$. Solving equations (1), (2), and (3a) for dC_P under this assumption one obtains

$$dC_P = \frac{-\lambda[1/(1+r)]^2}{M}\,.dt,$$

where
$$M = \begin{vmatrix} u_{11} & u_{12} & 1 \\ u_{21} & u_{22} & [1/(1+r)] \\ 1 & [1/(1+r)] & 0 \end{vmatrix}$$

Second-order conditions (strict quasi-concavity of the utility function) require that M be positive. Hence, we conclude that dC_P is negative. *Q.E.D.*

[1] t represents a temporary tax. Hence, if t is *not* zero this means that such a tax *already exists* and we are analysing the effects of increasing it However, the impact of any already existing taxes may be assumed to be incorporated in r, the real rate of interest.

THE STOCK MARKET VALUATION OF BRITISH COMPANIES AND THE COST OF CAPITAL 1955–69

By ANDREW GLYN[1]

SECTION I of this paper describes the results of estimating a model explaining the stock market valuation of a sample of British firms for the period 1955–69 From these results estimates are derived of the effect of various types of borrowing on market valuation (section II) and of the effect of dividend pay-out policy on valuation (section III) These estimates in turn provide the basis on which the costs of various types of capital are calculated (section IV), that is the rate of profit which must be earned on investment if it is to increase the stock market valuation of the shareholders' wealth The fact that estimates are made for each year allows an assessment of the influence of changes in the tax system on the cost for shareholders of retaining profits and of using various types of external finance The rate of profit declined over the period, and it appears that by the late sixties it was hardly, if at all, above the average cost of capital.

The data

Standardized Board of Trade company account data for 1954–5 to 1963–4 were available for 106 quoted companies with assets in excess of £2½ million in 1957, in the food, chemical, and engineering industries The short run of years ruled out any time series analysis of the valuation of individual firms; and even when additional data were secured for the following five years it was decided to confine the analysis to cross-sections Substantial changes in the structure of the valuation equation could be expected over the fifteen-year period and this, together with relatively small changes in the policy variables for most individual firms, would prevent useful estimates being derived from time series

Data on the market valuation of the firms had to be collected to complement the accounting data All firms were valued on the same day so that comparisons should not be distorted by changes in the general level of stock market prices Company account data used to 'explain' valuation should be the most recent available to the market This minimizes measurement errors when using accounting data as proxies for the stock market's expectations of future performance So firms should be included in the sample only if they had not published their interim dividend and

[1] I am most grateful to Bob Bacon, Anne Black, John Blandford, John Flemming, Peter Reeve, and David Soskice for help and advice while I was working on this material and particularly to John Wright for his patient supervision

report by the day on which valuations were taken, as this information would make the previous year's results much more out of date The majority of companies' accounting years end in December or March, the accounts were published typically three or four months after this, and the interim report came out some five months later By taking a day in the middle of July as the 'valuation day' it was possible to secure a sample of just fifty firms which, in each year of the period 1955 to 1964, had published their previous years' accounts and had *not* yet published their interim dividend for the current year The firms are listed in the appendix; when extending the data on to 1969 mergers took their toll of the sample so that by 1969 there were only thirty-five left

Ordinary shares, preference shares, and debentures were valued at the average of the buy-and-sell price as reported in the Stock Exchange Daily Official List for 20 July (or the next week-day) Numbers of the various types of securities were extracted from Extel Statistical Services Cards and were multiplied by their prices to give valuation The value of issues taking place between the end of the accounting year and 20 July was deducted from the over-all valuation since the earning power of the assets acquired with such issues would not be reflected in the previous year's accounts All security prices were taken 'ex dividend' but the post-tax value of dividends or interest was added back to security values Otherwise companies paying out high dividends would appear less highly valued, simply because a larger part of the current year's profits had been transferred to the shareholders and was thus excluded from the stock market valuation of the company

I. The basic model

The basic model used to explain the market valuation of the firm is an extension of that used by Modigliani and Miller (MM, 1966), and is shown in the following equation.

$$V = a_0 \log \text{ASSETS. DIST}^* + a_1 \text{DIST}^* + a_2 \overline{\text{RET}} + a_3 \text{PREF} + a_4 \text{DEB} + a_5 \text{BANK} + a_6 \text{TRADE} + a_7 \overline{\text{EXT}} + a_8 Z + u$$

where

V	= total market valuation of the firm = S+PREF+DEB+ +BANK+TRADE,
S	market valuation of equity,
PREF	market valuation of preference stock (PREF′ is book valuation),
DEB	market valuation of long-term debt (DEB′ is book valuation),
BANK	book value of bank overdrafts and borrowing,

TRADE	book value of trade and other credit received;
ASSETS	book value of assets $= E + \text{PREF}' + \text{DEB}' + \text{TRADE} + {} + \text{BANK}$;
E	book valuation of equity,
DIST*	total distributions (dividend plus interest) with interest scaled down to dividend equivalence;
$\overline{\text{RET}}$	a weighted average of the current and previous two years' levels of retentions; when reduced to dividend equivalence $= \overline{\text{RET}}^*$;
$\overline{\text{EXT}}$	an average of funds received by issue of equity, preference stock, long-term debt, bank borrowing, and trade borrowing over the current and previous four years;
Z	$(\overline{\text{EXT}}+\overline{\text{RET}}) \times \dfrac{(\text{DIST}^*+\overline{\text{RET}}^*)}{\text{ASSETS}}$,
u	a disturbance assumed to be independent of the explanatory variables and to be proportional to the level of DIST* for the individual firm.

These variables are as described by Singh and Whittington, some detailed points on definitions are made in the Appendix.

(a) Dependent variable

Most of the tests were run using valuation as the dependent variable, rather than dividend or earnings yield As MM (1966) point out, such 'yield-form' models involve serious problems of bias since V or S appears in the denominator of explanatory variables, making them dependent on the disturbance term in the equation

Levels of the different types of debt are included both as explanatory variables and as components of total market valuation (V) which is to be explained Correlation between the levels of debt and the disturbance term u would occur if there were random disturbances, or measurement errors, in the valuation of the different types of debt and this would bias down the estimate of the debt coefficients But this bias would not be affected by subtracting the debt variables from both sides of the equation so making the valuation of equity the dependent variable. For the disturbance term would be unaltered, as would its correlation with the debt variables still included as explanatory variables. Since it makes no difference to the estimates and simplifies their interpretation, V was used as dependent variable.

(b) Distributions, retentions, and growth

Distributed and undistributed profits are included separately in the equation in order to assess their relative effect on valuation Distributed

profits includes dividends on preference shares, and interest on debentures and bank-borrowing, as well as dividends on ordinary shares. This makes the split between distributed and undistributed profits dependent only on total profits and retentions, and independent of the levels of different types of debt. Thus in the model, gearing merely determines how distributed profits are allocated between ordinary dividends and other distributions, leaving retained profits and thus future internally financed growth unaffected This provides the best test of the alternative theoretical positions on gearing. For these suppose a given time stream of returns to the owners of the firm (holders of debt and equity), which may or may not be valued differently depending on how the stream is split between ordinary dividends and other distributions

The position is further complicated by the fact that interest payments were deductible for computing profits tax, and after 1964 corporation tax, liability. Thus £1 of extra interest would (assuming retentions constant) reduce dividends by less than £1 This means that extra gearing, even for a firm which kept retentions constant, would in fact be reflected in a level of distributions higher by the value of the tax savings on the interest payments But it is much simpler to have all the effect of gearing on valuation, including the valuation of the tax savings, reflected in the coefficients of the debt variables This can be achieved by scaling down the interest component of distributions to 'dividend equivalence' by deducting the tax savings. This adjusted distributions variable is then independent of the levels of the various types of debt

It has long been recognized that a major problem in estimating valuation equations is that cyclical and random disturbances will make actual retentions a very imperfect measure of the average level of retentions expected to accrue to the firm from the profits earned on its existing assets. Substantial measurement errors will bias down the coefficient of the retentions variable Also, given the likelihood of strong positive correlation over a sample of firms between distributions and expected retentions, as more profitable firms both retain and pay out more, the coefficient of the distributions variable will also be biased up. These biases obscure the relative importance of retentions and distributions in determining valuation. In the basic model the retentions variable is a simple weighted average of current and the two previous values (with weights of 0·6, 0·25, and 0·15 respectively); alternative approaches are outlined later on.

The growth of distributions accruing to the existing owners of the firm obviously plays a crucial role in determining valuation To include only a retentions variable would be a very crude way of capturing expectations about future growth since variations in the expected rate of profit on this

investment would be ignored. So would externally financed growth, which will contribute to the growth of distributions for existing owners if the external finance earns more than the cost of capital Taking the past use of external finance as a proxy for the expected future use, and the over-all rate of profit for the firm as determining expectations about the profitability of new investment, the variable Z is constructed It is an attempt to measure expected returns from internally and externally financed investment as the product of the expected amount of investment $(\overline{\text{RET}}+\overline{\text{EXT}})$ and the expected rate of profit $\frac{(\text{DIST}^{*}+\overline{\text{RET}})}{\text{ASSETS}}$.

There are, of course, well-known problems about the measurement of profitability from the rate of profit on book value, in particular estimation of depreciation and the book value of assets at historic cost will cause systematic biases in the measured rate of profit between companies with capital stocks of different average ages. There is no obvious way round this problem and the effect of such errors in variables just has to be faced. In measuring the rate of profit, retentions, as well as distributions, are measured in terms of 'dividend equivalence'; this involves deducting from actual retentions the extra tax (profits tax on distributed profits prior to 1958, and income tax after 1964) due had the retained profits actually been distributed. The measured rate of profit then depends only on profitability of the assets and not on gearing and dividend policies If these policies were reflected in the measure of profitability used the task of unscrambling estimates of the effects of these policies on valuation would be more complicated

A five-year average of the current and past rates of use of external finance is taken to measure expected future use of external finance A medium-term average is used in order to avoid distortions resulting from the large-scale and infrequent recourse to long-term external finance, and cyclical influences on the use of short-term finance The external finance variable is included separately in the model Its coefficient should measure the costs of using the finance (in terms of dividends and interest paid to the suppliers), while the benefits will be incorporated in the Z variable Even though the cost of retaining profits is already implicitly included in the equation in the form of distributions forgone, the retentions variable is also included separately. This is because retentions may have some effect on valuation quite apart from the expected returns from their reinvestment (for example the cushioning effect of retentions on dividends when profits fluctuate)

(c) *The debt variables*

Variables measuring the four main types of borrowing are included separately in order to assess their different effects on valuation

(d) The size variable and heteroscedasticity

It is widely believed that large firms are valued proportionately more highly than small ones Plausible reasons are that size is an indicator of diversification, and thus the stability of earnings, and also that a large volume of outstanding securities increases their marketability To capture any effect of valuation increasing more than in proportion, when both size and returns increase, a variable multiplicative in size (log ASSETS) and returns (DIST*) is included

With valuation of the largest firm more than 100 times greater than that of the smallest, the model as specified above is most unlikely to fulfil the requirement for Ordinary Least Squares to be the most efficient method of estimation, that the variance of the disturbance term should be equal for all observations Errors in explaining valuation would be expected to be roughly proportional to the size of the firm Tests of this assumption were made, using the procedure suggested by Park (1966), and, as expected, the standard deviation of the disturbance term was found to be approximately proportional to the size of the firm Tests suggested that deflating by DIST*, the single most important explanatory variable, would satisfactorily remove the heteroscedasticity and this procedure was used in estimating the basic model In certain later results deflation by ASSETS was more convenient, but prior tests showed that the use of a different deflator made negligible difference to the coefficients

(e) Estimation of the basic model

Table I presents the results of estimating the basic model by Ordinary Least Squares for the fifteen years 1955–69 The year refers to the July during which the securities making up V were valued

The effect of dividing through by DIST* to remove the heteroscedasticity is that $\bar{R}^2$ measures the proportion, adjusted for degrees of freedom, of the total variation in V/DIST* about its mean accounted for by the variation in the explanatory variables The constant term in the deflated equation, DIST* in the undeflated form, does not contribute to the explained sum of squares The appropriate F test indicated that the multiple correlation coefficient is significant at the 1 per cent level for each of the fifteen years, though only just so in 1960 The $\bar{R}^2$'s show a very definite pattern, before 1960 and after 1964 it is always greater than 0 5 whereas in the period 1960–4 it is always less than 0 5 This decline in the degree of explanation in the middle of the period cannot be explained by greater dispersion of valuation in those years as a result of a larger variance of the true disturbance term or of omitted variables If anything, the total sum of squares was at its lowest in the middle of the period, perhaps because of a tendency for the dispersion of pay-out ratios to

TABLE I

Regression estimates for basic model

Coefficients of	*1955*	*1956*	*1957*	*1958*	*1959*	*1960*	*1961*	*1962*	*1963*	*1964*	*1965*	*1966*	*1967*	*1968*	*1969*
DIST* log ASSETS	3 1	2 4	1 6	0 5	0 4	2 7	1 5	2 2	1 7	1 0	1 0	1 3	0 1	3 0	2 6
	(3 3)	*(3 9)*	*(2 0)*	*(0 7)*	*(0 4)*	*(2 8)*	*(1 2)*	*(2 1)*	*(1 7)*	*(1 2)*	*(1 3)*	*(2 4)*	*(0 1)*	*(1 5)*	*(1 6)*
DIST*	6 4	3 2	5 7	18 4	21 4	9 9	20 9	4 6	13 9	23 2	17 0	12 2	28 2	9 1	−3 0
$\overline{RET}$	2 9	2 4	0 3	1 6	9 5	−2 3	−7 8	−11 9	3 4	5 4	3 1	2 8	−2 4	−7 5	−3 9
	(1 1)	*(1 3)*	*(0 1)*	*(0 6)*	*(2 3)*	*(0 5)*	*(2 1)*	*(2 7)*	*(0 8)*	*(1 6)*	*(0 7)*	*(1 3)*	*(0 8)*	*(1 0)*	*(0 6)*
PREF	−0 38	−0 19	−0 0	−0 13	−0 16	−0 24	0 01	0 24	−1 03	−1 16	−0 78	0 04	−0 48	−1 72	0 80
	(1 4)	*(0 9)*	*(0 0)*	*(0 5)*	*(0 3)*	*(0 4)*	*(0 0)*	*(0 4)*	*(2 0)*	*(2 5)*	*(1 8)*	*(0 1)*	*(1 1)*	*(1 6)*	*(0 7)*
DEB	0 34	0 85	0 83	0 65	0 41	−0 15	−0 26	0 58	−0 11	0 19	0 48	−0 20	0 45	0 85	−0 11
	(1 1)	*(3 3)*	*(2 7)*	*(2 5)*	*(1 1)*	*(0 3)*	*(0 4)*	*(1 1)*	*(0 3)*	*(0 5)*	*(1 4)*	*(0 9)*	*(1 5)*	*(1 1)*	*(0 2)*
BANK	−0 30	1 09	0 56	0 32	−1 22	1 37	0 98	0 74	0 17	0 16	0 27	1 27	1 06	1 79	0 95
	(0 6)	*(2 9)*	*(0 7)*	*(0 6)*	*(1 2)*	*(2 3)*	*(0 9)*	*(0 7)*	*(0 5)*	*(0 3)*	*(0 4)*	*(3 5)*	*(2 8)*	*(1 4)*	*(0 7)*
TRADE	1 01	0 54	1 33	1 22	1 31	0 37	1 06	1 81	1 49	0 80	1 19	0 91	1 33	1 37	1 30
	(2 9)	*(2 3)*	*(3 7)*	*(2 9)*	*(1 9)*	*(0 5)*	*(1 3)*	*(2 8)*	*(3 8)*	*(2 5)*	*(4 9)*	*(5 0)*	*(5 7)*	*(3 5)*	*(3 3)*
$\overline{EXT}$	0 5	0 1	−1 5	−0 5	−0 3	−3 7	−6 8	−10 4	−2 7	−0 5	−2 6	−1 9	−5 7	−12 0	−5 8
	(0 4)	*(0 4)*	*(1 4)*	*(0 3)*	*(1 3)*	*(1 7)*	*(3 2)*	*(1 4)*	*(1 3)*	*(0 3)*	*(1 3)*	*(1 5)*	*(2 8)*	*(2 8)*	*(2 2)*
Z	24	27	78	30	−6	69	121	213	75	21	38	46	135	313	231
	(1 2)	*(1 8)*	*(3 9)*	*(1 0)*	*(0 6)*	*(1 7)*	*(3 5)*	*(4 6)*	*(2 0)*	*(0 8)*	*(1 1)*	*(1 9)*	*(3 2)*	*(3 2)*	*(3 5)*
R^2	0 71	0 81	0 81	0 61	0 52	0 25	0 35	0 43	0 47	0 35	0 52	0 87	0 85	0 52	0 78
RSSQ	1602	640	1269	980	2347	1802	3181	2324	1923	1494	1073	455	875	4135	2013
TSSQ	4951	4059	7883	3012	5813	2887	5838	4909	4369	2756	2052	4200	7262	10785	11894
Mean V/DIST*	52 1	40 8	44 3	39 2	45 0	42 2	43 2	39 8	43 9	43 0	38 0	40 0	42 9	56 2	47 5

t values are in brackets

decrease after the ending of the differential profits tax. The failure of the included variables to explain as much variation in V/DIST* in the middle of the period is surprising since it might have been imagined that increasing sophistication of investment analysis in the early sixties would have resulted in more 'objective' criteria for equity valuation. Nor is there any evidence that the degree of explanation varies with the over-all level of market optimism or pessimism, at least as measured by the mean of V/DIST* Although the variation in the explanatory power of the model is rather peculiar, to score $\bar{R}^2$'s of more than 0·75 in five out of fifteen cross-sections is a definite achievement

II. The effects of debt on market valuation

With the levels of assets, retentions, and tax-adjusted distributions held constant in the regression, the effect of the various types of debt on market valuation of otherwise identical firms can simply be read off from their respective coefficients

(a) *Preference stock*

a_3 the coefficient of the preference stock variable is negative in eleven out of the fifteen years, but is only significant at the 5 per cent level in 1963 and 1964 when its value is close to -1 (implying that the total market valuation of a company was lower roughly in proportion to the value of preference stock outstanding) Under the profits and corporation tax systems preference dividends (unlike interest) attracted no tax concessions This means that the predicted value of a_3 is zero on MM's hypothesis of the irrelevance (apart from tax savings) of gearing for total valuation Only in two years is a_3 significantly different from zero, though the persistence of the negative sign, and the average value of $-0{\cdot}35$, suggests some damaging effect from the use of preference capital, particularly in the sixties With the share of preference capital in the book value of total capital (long and short-term) of quoted manufacturing companies falling by about a half between 1953 and 1960 it is hard to believe that gearing with preference capital was being carried to 'excessive' lengths, which would be the explanation for the negative coefficients on 'traditional' hypothesis of the effect of gearing on market valuation One possibility is that, with the substantial tax advantages offered by borrowing by means of debentures, it was only the conservative, sluggish companies which still had a substantial part of their capital in the form of preference shares, so that a high level of preference capital acts as an indicator of poor growth prospects In fact the use of preference capital is slightly positively correlated both with profitability and (more to the point since

it is not included in the valuation model) the stock of liquid assets which might be an indicator of conservatism.

For the years 1955–64 attempts were also made to assess any non-linearity in the relationship between V and PREF by adding a variable multiplicative in PREF and the ratio of the book value of preference stock to the book value of equity In the first half of the period the relationship was of the form predicted by the traditional hypothesis, with the use of preference capital typically increasing V up to the point where it was 15 per cent of the book value of equity (the sample average ratio in 1957 was 12 per cent) In only one year, 1959, was the relationship significant, with preference stock decreasing V to a smaller and smaller extent as its importance increased [1]

Least squares bias would occur in these estimates if a high level of valuation encourages use of preference capital, perhaps because highly valued firms could more easily raise the capital [2] But estimates with the preference variable replaced by an instrument, derived from regressing PREF on profitability, industry dummies and liquidity, yielded coefficients which were on average little different from the OLS estimates

(*b*) *Debentures*

The coefficient a_4 of the long-term debt variable is positive in ten out of the fifteen years, but only significant during the period 1956–8 The average value of the coefficient is 0 32, which is very close to the average value of tax savings generated by debt interest (0·30 of debt interest) On MM's hypothesis the coefficient of the debenture variable would be equal to the rate of tax savings (see MM (1966), p 340) and possibly greater if these tax savings were expected to grow. On only two occasions is the estimate for a_4 significantly different from that predicted by the MM hypothesis—in 1956 the coefficient is significantly more positive whereas in 1966 it is significantly less positive. The latter case could perhaps be explained by the problems for the market of interpreting accounts for 1965 as they were distorted by the transitional provisions of the 1965 Finance Act Over all, however, the coefficients do not closely follow the trend in tax savings as the following table shows; though given the insignificance of the coefficients perhaps too much should not be read into this:

[1] Tests using the ratio of preference capital to a weighted average of distributions and retentions, as a proxy for the ratio of the market value of preference stock to V, yielded totally insignificant results

[2] This is on the assumption that there are persistent 'firm effects' reflected in the disturbance term, so that there is a correlation between past ability to raise debt and the current disturbance In fact for the years 1955–64 more than half the residual sum of squares was accounted for by persistent firm effects.

TABLE (a)

	1955–8	*1959–61*	*1962–5*	*1966–9*
Debenture coefficients	0 67	0 00	0 28	0 25
Tax savings per unit of debt interest	0 32	0 18	0 26	0 42

Tests for non-linearity in the relationship between V and DEB indicated a persistent tendency for DEB to lower V initially—typically until the ratio of debentures to the book value of equity was around 12 per cent (about the sample average)—but that after this level further borrowing would raise market valuation In only three years were the coefficients significant, but still the non-linearity is just the opposite of that predicted by the 'traditional' hypothesis according to which further borrowing would begin to reduce valuation only when it had become 'excessive'

Replacing the actual value of DEB by an instrument yielded estimates which were on average very close to the OLS estimates. Significance levels were much reduced, which is not surprising since the correlation between DEB and the instrument was slightly less than 0·5 Firms with a high ratio of debentures to total assets appeared to be of less than average profitability and with a high proportion of fixed assets against which debentures could be secured

(c) *Bank borrowing*

The coefficient of BANK (a_5) is positive in every year except two, and is four times significant The average value of a_5 is 0·64, about double the average rate of tax savings (which are the same as in the case of debentures), though the estimated coefficient is significantly different from that predicted by the MM hypothesis (the rate of tax savings) only in 1956, 1959, and 1966 The difficulty of distinguishing between competing hypotheses is well shown by the fact that the coefficient is also only twice significantly different from 1. It is of some interest that all the years in which a_5 is really large are in periods of generally tight money (1955–7, 1960–1, and 1966–9) Again this might be explained by a tendency for the most credit-worthy, and therefore more highly valued, firms to have greater access to bank borrowing in tight money periods.

On three occasions there appeared to be a significant non-linear relationship between BANK and V, but in all cases borrowing from the bank appeared to have a decreasingly adverse effect. In 1963, for example, when the relationship was most significant, bank borrowing only increased

valuation after it reached a ratio to equity assets of about 10 per cent, about twice the sample average.

About 60 per cent of the sample variation in the ratio of bank borrowing to total assets was accounted for in terms of variations in profitability, industry group, and structure of assets It is correlated positively with the extension of trade credit and negatively with holdings of liquid assets Using an instrument for BANK yielded an average estimate for a_5 of 0·4 for the years 1955–64, as compared with 0 6 for comparable OLS estimates, suggesting that some bias may have occurred in the OLS estimates as a result of more highly valued firms finding it easier to borrow from banks

(*d*) *Trade credit*

The coefficient of the trade credit received variable, a_6, is significant far more often than any of the other debt variables—twelve times at the 5 per cent level The average value of this coefficient is 1·1 Only in 1956 is a_6 significantly different from 1 at the 5 per cent level A possible explanation for the high estimated value for a_6 is the following The amount of trade credit received by a company may notionally be split into an interest-free tranche where taking credit incurs no costs, and a discounts-lost tranche where the credit has an interest cost dependent on the rate of loss of discounts for quick payment for purchases If a company is expected to retain at least the current level of interest-free trade credit indefinitely (not necessarily from the same supplier), then it would represent no claim on the income or assets of the company. Trade credit would have no market value *per se*, the supplier of the credit would not be able to sell this liability of the company for anything So the value of trade credit received in the interest-free tranche should not be included in the total market valuation of the firm Its inclusion as part of valuation would lead to an estimated coefficient of 1 provided there was no genuine effect of this costless trade credit on true total market valuation

The argument is rather less precise for the discounts-lost tranche The value of these discounts lost cannot be measured and included in profits; this means that a_6 measures the valuation of sufficient profits to pay for the discounts lost. Even if the effective rate of interest was so high that a proper market valuation of this liability would be much greater than the book valuation, it is the book valuation which is included in V Thus the coefficient of TRADE would be 1 if extra trade credit in the discounts-lost tranche had no effect on V It might be less than 1 because of the effect on the stability of the returns to owners of other types of capital if there were demands for repayment of trade credit.

There is a persistent, though generally insignificant, tendency for the

positive effect of trade credit on valuation to decline with the level of credit received and to become negative beyond the point where the ratio of trade credit received to equity assets was around one-third (compared with a sample average value of about one-fifth). This effect is in accordance with the 'traditional' hypothesis that gearing has a decreasingly beneficial effect on valuation; the weakness of the relationship may be due to multicollinearity between the two trade-credit variables included in the tests for non-linearity.

There was quite a strong positive association of trade credit received with profitability and with the extension of trade credit, and a negative association with holdings of liquid assets, in all, more than two-thirds of the variations in trade credit could be explained The estimates of a_6 using an instrument for TRADE yielded an almost identical average for the coefficients as when OLS was used, suggesting that bias is probably not a serious problem.

(e) *Other tests relevant to hypotheses about gearing*

(i) *Risk* The basic model contains no variables measuring the riskiness of the returns to the individual companies and this omission might bias debt coefficients if risky, and therefore less highly valued, firms raised less debt The most usual way of measuring risk is to take something like the standard deviation of earnings about their trend. When such a variable was added to the basic model it had the wrong sign more often than not, and was only significant on one occasion and then it had the wrong sign.[1] This confirmed *a priori* scepticism about whether, given the policy of stabilizing dividends, this dimension of risk would be of any significance for market valuation. What seemed a more serious type of uncertainty was not variability of earnings about its trend, but rather uncertainty about the trend itself This is obviously extremely difficult to measure, but an attempt to do so was made by including a variable measuring the maximum percentage fall in cash flow (DIST+RET+DEP(reciation)) from peak to trough, suffered by the company since 1951. The coefficient of this variable had the right sign in every year It was significant in the years 1960–2, by which time it is reasonable to suppose the variable captures *uncertainty* about the trend, rather than the depressing effect on expectations of a substantial recent fall in cash flow. There were no substantial effects on the debt coefficients when these risk variables were included.

(ii) *Industry effects.* The basic model was also estimated (after deflation by DIST*) with dummies included for the food, chemicals, and electrical

[1] The variable used was the coefficient of variation of the rate of profit (before deducting depreciation which might otherwise introduce spurious fluctuations) over a five-year period.

engineering industries. On three occasions all three dummies were positive, significant, and of the same order of magnitude, in the years 1963, 1967, and 1968 the ratio V/DIST^* was on average some 7, 4, and 15 lower for a firm in the non-electrical engineering than for a firm of equal profitability in the other industries. In 1968 this difference was about one-quarter of the mean value of V/DIST^* Despite these very substantial effects the inclusion of these industry dummies did not materially affect the debt coefficients

(iii) *Yield-form models* Some experiments were made for the years 1955–64 using a 'yield-form' model with DIST^*/V as the dependent variable—a similar dependent variable to that used by Davenport (1971) except that here interest is reduced to dividend equivalence Ratios of the book value of the various types of debt to the book value of equity were used to measure the different types of leverage, and the past rate of growth of assets used as a measure for the expected rate of growth. The following table compares the average effect of the various types of debt on valuation estimated from the basic and yield-form models:

TABLE (*b*)

	Average effect on valuation of an extra £ of			
	PREF	DEB	BANK	TRADE
Basic model	−0 31	0 33	0 42	1 09
Yield form	−0 11	0 47	1 11	1 00

The yield-form estimates suggest rather more positive effects of debt on valuation, particularly in the case of BANK which has generally more significant coefficients as well Tests for the non-linearity of the relationship between levels of debt and valuation yielded a thoroughly confused picture Only PREF consistently accords with the 'traditional' hypothesis that gearing initially increases V, but then has an adverse effect; DEB also has this effect early in the period in contrast to the results on the basic model Moreover, unlike the results for the basic model, there was no tendency for TRADE to have a decreasingly advantageous effect on valuation

(*f*) *Summary and comparison with Davenport's results*

These tests of the effect of leverage on valuation do not yield decisive support for the traditional view, or MM s or indeed any other. As far as long-term debt is concerned there is some weak evidence for the traditional view that gearing with debt or preference stock has a decreasingly beneficial effect on market valuation, though it is really not possible to reject

with any confidence the MM hypothesis that the only effect is through the tax concession on debt interest. Certainly these results provide less firm support for the traditional view than do those of Davenport. Two aspects of Davenport's model are worth commenting on here In common with most other work, he aggregates preference stock and long-term debt when deriving his gearing variable The results for the basic model discussed above at least suggest that these two types of debt may have rather different effects, and not just because of the tax concession to debt, in which case such aggregation may be invalid Secondly, though his samples are generally larger and more homogeneous than mine, his level of explanation is less high Using his model I the highest corrected correlation coefficient obtained is 0·37—for the same three years this is less than the lowest $\bar{R}^2$ on my roughly comparable yield-form model He uses the past rate of growth of earnings, but short-run fluctuations may make this a worse indicator of expected growth than the rate of growth of assets, certainly the rate of growth of assets performed much better in my sample

The results for short-term debt are a bit clearer Trade credit probably increases total market valuation proportionately up to the discounts-lost tranche, and the model suggests that there may be decreasing beneficial effects thereafter. Bank credit certainly increases total valuation by the amount of the tax savings generated; the larger coefficients which occur in times of monetary restraint may perhaps be explained by some correlation of bank borrowing with safety rather than genuine effect of bank borrowing *per se* on valuation.

III. Effect of distributions, retentions, and external finance

This section presents estimates, derived from the coefficients of Table I, of the valuation of distributions and retentions Finally, the effect of substituting distributions for retentions is considered

(a) *The effect of distributions*

The effect of an additional unit of distributions on valuation, with everything else (RET, EXT, ASSETS, etc.) held constant, is found by taking the first derivative of V with respect to DIST* and evaluating the resulting impact multiplier:

$$M_{\text{DIST}^*} = a_0 \log \text{ASSETS} + a_1 + a_8(\overline{\text{EXT}} + \overline{\text{RET}})/\text{ASSETS}$$

The total effect on valuation comprises the constant a_1 and factors dependent on the size of the firm and the rate of growth of the company's assets The rate of growth determines the importance for shareholders of the increase in the rate of profit earned by the company implied by an additional unit of distributions.

a_0 is always positive and is significant at the 5 per cent level on six occasions, but fluctuates quite widely. a_1 is also always positive and fluctuates in the opposite direction to a_0. In the deflated form of the basic model used to estimate the coefficients of Table I, a_1 is the constant term and there was no facility on the regression programme for testing its significance. a_8 is positive in every year except 1959, is significant on seven occasions, and has regular cyclical peaks in 1957 and 1962 and 1968. Given the inevitable multicollinearity that results from the construction of Z out of the variables DIST*, $\overline{\text{RET}}$, and $\overline{\text{EXT}}$, which also appear separately in the model, this level of significance for a_8 is rather satisfactory.

The following table gives the estimated values of M_{DIST^*} for a firm with sample average values for $\overline{\text{EXT}}$, PROF/ASSETS and RET/ASSETS, and with ASSETS of £5 million—which is near the median for the sample.

TABLE (c)

The valuation by the stock market of an extra £ of distributions

	(£)		(£)
1955	35	1963	33
56	26	64	33
57	25	65	28
58	25	66	26
59	25	67	35
60	37	68	51
61	41	69	33
62	37		

There is a good deal of year fluctuation in the valuation of an extra £ of distributions, with peaks usually corresponding more or less with peaks in V/DIST*. The valuation put on the distributions of giant firms (assets of £500 million) is on average slightly less than 20 per cent greater than for firms with assets of £5 million, with the effect of size varying with a_0 in Table I. Greater use of external or internal finance by a firm increases M_{DIST^*} to an extent depending on the valuation of future growth, but only in the years 1962, 1968, and 1969 (when a_8 is really large) does the effect of having values of $\overline{\text{RET}}$/ASSETS or $\overline{\text{EXT}}$/ASSETS of half as much again as sample average increase the estimate of M_{DIST^*} by as much as 10 per cent.

(b) The effect of retentions

The multiplier for an extra unit of retentions (*not* scaled down to dividend equivalence) is given by:

$$M_{\text{RET}} = a_2 + a_8 t(\overline{\text{EXT}} + \overline{\text{RET}})/\text{ASSETS} + a_8(\overline{\text{RET}}^* + \text{DIST}^*)/\text{ASSETS}$$

The effect of an extra unit of retentions depends both on profitability (of the investment financed by the retentions) and the average rate of investment (the profitability of which is increased if retentions are higher but distributions unchanged). When a_8 is large, a_2 tends to be negative (though only twice significant) which strictly implies that retentions yield no benefit until investment and profitability reach certain levels; when a_2 is positive, retentions yield some benefit regardless of the rate of return earned by the firm. Since $\overline{\text{RET}}$ enters twice into the construction of Z there is a severe multicollinearity between the two variables, with the correlation between them often being around 0·8, and so the estimates for a_2 and a_8 must be treated with caution

The following table gives the estimated values of M_{RET} for a firm with sample average values for relevant variables

TABLE (*d*)

The valuation by the stock market of an extra £ of retentions

	(£)		(£)
1955	6	1963	12
56	6	64	8
57	9	65	8
58	5	66	6
59	9	67	7
60	7	68	14
61	9	69	13
62	16		

Again there are quite substantial fluctuations from year to year A rate of profit half as high again as the sample average increases M_{RET} by the order of one half in 1957, 1960–2, and 1967–9 (years with high values of a_8) This is because retentions are valued more highly in firms with a high rate of profit.

(*c*) *The effect of external finance*

The impact multiplier for an additional unit of external finance is given by $M_{\text{EXT}} = a_7 + a_8(\overline{\text{RET}}^* + \text{DIST}^*)/\text{ASSETS}$ The second term takes account of the impact of the current rate of profit on the benefit from employing external finance a_7 is negative except for the first two years, but it is substantial and significant only in 1961–2, and 1967–9, years when a_8 is large and positive Predominantly negative signs accord with the interpretation that a_7 is the product of a (negative) coefficient measuring the generation of costs by the use of external finance and a (positive) coefficient representing their capitalization Estimates for M_{EXT} on the

alternative assumptions of average and high profitability (half as much again as the average) are shown below:

TABLE (e)

The valuation by the stock market of an extra £ of external finance

Year	Profitability Average (£)	Profitability High (£)	Year	Profitability Average (£)	Profitability High (£)
1955	2	2	1963	1	4
56	2	2	64	1	1
57	3	6	65	0	1
58	1	2	66	0	1
59	−1	−1	67	0	2
60	1	3	68	−1	5
61	1	5	69	3	7
62	2	8			

For firms earning an average rate of profit the use in the past of an additional £ of external finance usually increased valuation a bit. For firms earning a high rate of profit the effect on valuation of past use of external finance was generally much greater, especially in the early and late sixties Presumably the costs of external finance are not greater for more profitable firms, provided their profitability is reflected in their share price, so that the extra returns from more profitable investment accrue to existing shareholders. The expectation of this occurring from future investment is reflected in the extra present valuation of the profitable firm which uses more external finance.

(*d*) *Increasing the pay-out ratio*

The effect of substituting a unit's distributions for retentions is derived by simply subtracting the multiplier for retentions from the multiplier for distributions With 'discrimination' against dividends at the beginning and end of the period, M_{DIST} must be scaled down by the appropriate tax factor which gives the value of extra dividends resulting from foregoing a £'s worth of retentions With levels of debt, and therefore interest payments, held constant in the model this gives estimates of the effect on valuation of substituting ordinary dividends for retentions In the years 1959–65, when there was no discrimination against dividends, increasing dividends, at the expense of £1's worth of retentions, would have increased a firm's valuation by more than £20. The effect, with the exception of 1955, is considerably smaller when higher tax rates on distributions were in force. At the end of the period particularly, the differential is somewhat smaller for more profitable firms, their retentions were valued more highly

TABLE (f)

Effect on valuation of forgoing £1 of retentions

	Average profitability $M_{t\,DIST^{*}-\overline{RET}}$ (£)	*High profitability* $M_{t\,DIST^{*}-\overline{RET}}$ (£)	*Dividends payable from £1 of retentions*
1955	20	19	0 74
56	13	12	0 72
57	8	6	0 68
58	12	11	0 68
59	16	16	1 00
60	30	28	1 00
61	32	28	1 00
62	21	15	1 00
63	20	18	1 00
64	25	24	1 00
65	20	18	1 00
66	7	6	0 59
67	14	11	0 59
68	16	9	0 59
69	6	0	0 59

because of being expected to yield higher profits But these results sugge that it is *only* the higher taxation of dividends which makes the bene from increasing the pay-out ratio lower in the late sixties than in the ear sixties. The extra valuation of a unit of dividends compared with a unit retentions is given below in Table (*g*) (ignoring the fact that the ta system prevented a one-for-one substitution before 1958 and after 1965)

TABLE (*g*)

Excess of stock market's valuation of £1 of dividends over valuation of £1 of retentions

	1955–8	*1959–62*	*1963–5*	*1967–9**
$M_{DIST^{*}-\overline{RET}}$ (£)	21	25	22	25
Sample rate of profit (%)	7 5	6 6	5 8	5 3

* 1966 omitted because of transition provisions of 1965 Finance Act Figures f 1967–9 include investment grants

It can be seen that there is no downward trend in the relative valuation distributions such as might have been expected from the growing empha on 'growth stocks' and on evaluating shares by 'price-earnings' rati rather than dividend yields. But the table also shows that there was marked downward trend in the rate of profit earned by the sample of firn

(in common with U.K. firms as a whole). If achieved profit rates are taken as some guide to expected future profitability, then a *fall in the profit rate* would be expected to reduce the relative valuation of retentions through a dampening effect on expectations of the future profits accruing from their reinvestment So the fact that the relative valuation of dividends did *not* increase over the period suggests that more weight was being placed on future growth through retained profits, though falling profitability prevented this fact from increasing the *valuation* of retentions[1]

A further refinement in estimating the effect of substituting dividends for retentions is to assume an offsetting increase in external finance used so that past investment (and by assumption expected future investment) is held constant This approximates most closely to MM's hypotheses (see MM (1961)) as to the irrelevance of dividend policy for valuation All that has to be done is to add M_{EXT} from Table (*e*) to $M_{t\,\mathrm{DIST}^{*}-\overline{\mathrm{RET}}}$ in Table (*f*) Typically the small positive values for M_{EXT} slightly push up the estimated effects of the substitution Perhaps the surprising thing is the beneficial effect on valuation of the use of external finance, given the rather low valuation set on retentions If all investment was expected to be as unprofitable as the valuation of retentions would suggest, it seems likely that existing owners of the firm would be expected to suffer from the use of external finance carrying a definite cost in terms of interest or dividends One explanation for the beneficial effect might have been that much of the external finance raised had very low cost (trade credit for example), but firms differed more in the amount of new equity and debenture finance raised, and these sources of finance certainly have substantial costs

So the most likely explanation is that some companies which were expected to have really profitable investment opportunities, not reflected in their achieved profits rates, raised external finance, though several of the firms which raised a lot of external finance (for example, AEI and Vickers) do not seem to be in that category It does not seem to be the case that the estimated effect of external finance is biased up due to dependence on willingness and ability to raise external finance on valuation There was no very clear pattern in the effect on M_{EXT} when actual $\overline{\mathrm{EXT}}$ was replaced by an instrument

All this does not change the central conclusion from the basic model that, except when the tax system makes paying out of dividends extremely costly in terms of retentions forgone, firms which pay out more dividends are valued a very great deal more highly than firms which retain a higher proportion of their profits.

[1] The quantitative importance of retentions declined considerably—from almost 5 per cent of assets in 1955–8 (accounting years 1954–7) to less than 3 per cent in 1967–9

(*e*) ***Bias in the estimates***

It has already been pointed out (p 216) that errors in measuring expected retentions may seriously bias down the estimated effect of retentions on valuation and bias up the estimated effect of distributions That these biases could account for the apparent higher valuation of companies which pay out a large proportion of their profits should not be ruled out. The likelihood of this is greater when it is recognized that the estimation of the relative valuation of dividends and retentions is also beset by simultaneous equation problems Companies which were highly valued might feel obliged to pay out a higher proportion of their profits so that shareholders received an adequate income on their investment Again high stock market valuation may facilitate and encourage the use of external finance, leading to less pressure to retain profits for the financing of future investment. If for these reasons, the pay-out ratio *was* also dependent on valuation, as well as vice versa, then Ordinary Least Squares will lead to (upward) bias in the estimates of the effect of higher pay-out on valuation

MM (1966) attempted to deal with the problem of errors in the profits variable by constructing an instrument derived by regressing profits on dividends, the level and growth of assets, and some debt variables Their rationale was that the companies' policies of dividend stabilization meant that 'dividends and dividend changes indirectly convey a great deal of information about management's expectations of long-run profits' (p 354).

The overwhelming disadvantage of this approach is that dividends are likely to so dominate the instrument constructed for profits that it becomes impossible to test the relative effects on valuation of dividends and retentions. The obvious answer would be to construct instruments for both distributions and retentions But there are insufficient other variables which could plausibly be argued to be correlated with the true levels of distributions and retentions So it would be impossible to have much confidence in the results of such a procedure An alternative approach was tried which involved constructing an instrument for DIST*, by regressing it on profitability and other variables, and deriving an estimate for $\overline{\text{RET}}$ by subtracting the instrument for DIST* from actual profits This does not eliminate the problem of bias, but by feeding part of the measurement error in profits into the instrument for DIST*, rather than leaving it all to show up in actual retentions, the exaggeration of the *relative* valuation of distributions should be reduced

The results from this TSLS procedure did suggest that the OLS estimates may have exaggerated the relative importance of distributions at the beginning of the period, though in the middle any bias seemed very slight. But in four out of the last five years these alternative estimates showed retentions being far *more* highly valued than distributions—totally

reversing the OLS results It is possible that there was a substantial shif towards valuing retentions more highly at the end of the period which wa masked by biases in the OLS estimates Certainly the simple correlatio between $\overline{RET}$ and V was considerably below that between DIST* and I in 1963 (0·72 as compared with 0 86—all variables being deflated b ASSETS) but much higher by 1969 (0 79 as compared with 0·48) Bu my own inclination is to rely on the Table I OLS estimates as thes alternative estimates show quite implausible fluctuations in the relativ valuation of a £'s worth of dividends and retentions (from 27 11 in 196- to 9 13 in 1966, to 27 6 in 1967, to 7 25 in 1968)

One further source of upward bias in the coefficient of the distributior variable would come from correlations of (1) the current level of di tributions with their past rate of growth (firms with high pay-out ratic might have tended to have rapidly growing distributions) and (2) the pas rate of growth of distributions with future growth expectations nc *captured in the Z variable* But when Z was replaced by a variabl measuring the past growth of dividends the coefficient of the latter wa never positive and significant In fact the retentions variable used up t 1958 was derived by multiplying the weighted average of current and pas values of $\overline{RET}$ by the ratio of the current value of DIST* to a similarl weighted average of current and past values of DIST* The idea behin this adjustment is the familiar one of the information contained in dividen decisions—in a rough way the past values of retentions incorporated in tl weighted average are increased proportionately to any growth in dividen which occurred subsequently Up to 1958 this modification of the r tentions variable improved the model's explanation a bit as compared wit a simple weighted average, but in 1959 its use led to a drastic worsenin The explanation is that after the change in the company tax system 1958 companies were encouraged to pay out a higher proportion of the profits as dividends This destroyed, in the short-run, the reliability dividend changes as an indicator of expected profit changes So up 1959 the inclusion of past dividend changes did improve the model, bu this was adequately incorporated in the basic model Thereafter pa dividend changes were not significant, so that their omission could n conceivably lead to serious bias

The period 1965–9 offered some interesting possibilities for tryi retentions variables which reflected changes in, or different treatments c the tax system, but the results were disappointing Recalculating r tentions (in 1964 in particular) to take account of the tax system actual in force at the time of valuation, rather than that reflected in the account made negligible difference Subtracting the reduction in taxation resul ing from the receipt of investment allowances, on the grounds that the

reflected in part the future growth rather than the present profitability of the firm, made virtually no difference Adding investment grants received by companies to their retentions led to a slight increase in explanation of valuation in 1967, a marked rise in 1968, and a marked fall in 1969 Again investment grants in part reflect growth rather than existing profitability and on these grounds perhaps should be excluded from measured retentions They were quantitatively important, however, on average adding 50 per cent to retentions, and it is peculiar that the way they appeared to affect valuation was so different in 1968 and 1969

Finally, the yield-form model (see p 225) gave rather higher estimates for the valuation of distributions The average effect of substituting a unit's distributions for retention was to increase valuation by twenty-seven units for the period 1955–64, compared with twenty units on the basic model

IV. The cost of capital

The estimates for the valuation of distributions and retentions described earlier provide a basis on which the cost of capital can be calculated For they enable an estimate to be made of the *size* of the stream of distributions and retentions which must be earned on an investment if the increase in the valuation of the company is to be sufficiently large to justify raising the finance (see MM (1966), pp 336–48) The estimates for 1955 in Table I may be used to illustrate the procedure

Calling the rate of profit from a project r, and assuming a sample average pay-out ratio $\left\{\frac{\text{DIST*}}{\text{DIST*}+\text{RET}}\right\}$ of 0·38, a £1 project would yield an annual stream of distributions of £0·38r, and a stream of retentions of £0·62r each year The estimates of the effects of distributions and retentions on valuation, shown in Tables (*c*) and (*d*), are 36 and 6 respectively so the total valuation of the profits from the investment project would be £$36\times0\ 38r+6\times0{\cdot}62r$. An extra unit investment involves an increase in assets, this reduces the rate of profit (leaving aside the profits from the investment which have already been dealt with), thus lowering valuation, but also increases the size of the firm The coefficients in Table I allow the total effect of higher assets to be calculated at $-0{\cdot}06$ The gain to existing shareholders is the total increment to valuation [£$36\times0{\cdot}38r+6\times0{\cdot}62r-0{\cdot}06$] less the unit of finance newly subscribed If the investment-cum-financing package is to leave existing shareholders just indifferent, then the rate of profit which must be earned (the cost of capital) will make the net gain to existing shareholders equal zero [1]

[1] Since distributions are calculated net of income tax at the standard rate, the costs of capital must be interpreted with income tax deducted from distributions paid from the profits on the investment

If the investment is financed by increasing one of the forms of borrowing, rather than issuing shares, then the effect on valuation of an extra unit of the relevant type of debt must be taken into account. Thus with debentures having a coefficient in 1955 of +0·34 the cost of using debenture capital in 1955 is the solution to r in the following equation

$$36\times 0{\cdot}38r+6\times 0{\cdot}62r-0{\cdot}06+0{\cdot}34 = 1 \text{ (and } r \text{ turns out to be 4·0 per cent)}$$

If the investment is financed internally then the situation is the same as in the case of external equity finance except when there is tax discrimination against dividends When additional tax is paid on distributions, the cost to existing shareholders of retaining £1, rather than distributing it, is not £1 but somewhat less (£0 79 in 1955 for example) Correspondingly the required increment to valuation, which would leave the shareholders indifferent as between retentions and distributions is less (£0 79 in 1955). The temporary impact effect on valuation of the reduction in the pay-out ratio (see section II) implicit in the extra investment is ignored The valuation of additional profits in the next year is assumed to be decisive, at which stage the pay-out is not affected by the current decision to invest No assumption about the certainty of future profitability is required All that is necessary is that the profits from the proposed investment should be reckoned to be of similar riskiness to the profits from present operations, for this degree of riskiness is assumed in the estimates of the valuation of existing profits The calculation is of the rate of profit to be earned *immediately*[1] on the book value of the investment (i e the real rate of profit) so that the estimates for the cost of capital are also in real terms Any increase in money profits, and so the book rate of profit, which may be expected to result from inflation in the future is irrelevant since again its effect on valuation (if any) is embodied in the estimates of the valuation of existing profits

The figures in Table II bring together the estimates of the cost of various types of finance as derived from Table I coefficients Four-year averages are shown to illustrate trends, and to exclude temporary fluctuations which are probably of no real significance Other facts on financing, interest rates, and so on are also given in the table as they are helpful in interpreting the results

Starting with the cost of external equity finance (line 7), the estimates show something of a downward trend, hovering about 6 per cent since the late fifties This was well above the earnings yield (line 9) by 1966–9, and around double the dividend yield (line 10) This estimate of the cost of

[1] If the stream of real profits was not constant, this may cause problems for a firm which may see its valuation reduced if, for example, its projects have a long gestation period But in the long run it must earn the same rate of profit on such investments as on the simple type of perpetuity investment considered here.

Table II

Estimates of the cost of capital % p.a.

		1955–8	1959–62	1962–5	1966–9
Equity					
	Proportion of new finance from				
(1)	Internal equity (inc. depreciation)	62.6	57.6	57.0	44.6
(2)	External equity	13.7	18.6	9.9	18.9
(3)	Sample average pay-out ratio	0.42	0.47	0.55	0.49
	Estimated cost of				
(4)	Internal equity (average firm)	5.7	6.3	5.5	4.1
(5)	,, (giant firm)	4.8	5.3	4.7	3.5
(6)	,, ,, (high pay-out)	4.5	4.5	4.1	3.3
(7)	External ,, (average firm)	7.6	6.3	5.5	6.3
(8)	,, ,, (constant pay-out)	6.8	6.0	5.7	6.3
(9)	Earnings yield on equity	12.0	7.7	5.9	4.4
(10)	Dividend yield on equity	3.3	3.0	3.2	2.9
(11)	Real return on holding shares	2.8	11.8	−0.2	7.5
Debentures					
(12)	Proportion of new finance	8.7	6.6	10.5	14.9
(13)	Estimated cost	2.7	5.5	4.2	5.2
(14)	Real cost of interest	1.2	1.0	−0.4	−1.0
(15)	Real return on holding debentures	−6.7	−2.4	0.5	−6.6
Preference stock					
(16)	Proportion of new finance	1.5	0.9	0.9	−0.8
(17)	Estimated cost	8.5	6.4	8.9	8.4
(18)	Real cost of dividends	0.1	1.9	0.8	0.7
Bank borrowing					
(19)	Proportion of new finance	3.8	5.7	6.9	6.6
(20)	Estimated cost	3.6	4.8	3.9	3.2
(21)	Real cost of interest	−1.0	0.7	0.5	−1.2
Trade credit					
(22)	Proportion of new finance	9.7	10.7	15.0	15.8
(23)	Estimated cost	1.0	1.0	0.0	3.0
Averages					
(24)	Estimated average cost	5.3	5.7	4.5	4.5
(25)	Distributions yield (sample)	2.3	2.4	2.4	2.2
(26)	Overall earnings yield (sample)	5.5	5.1	4.4	4.5
(27)	Sample rate of profit	7.5	6.6	5.8	5.3

Sources

Lines (1), (2), (12), (16), (19), (22), *Economic Trends*, April 1962, Table 2c, and Business Monitor M3 1970, 1971. Coverage: quoted manufacturing companies.

Lines (9), (10), (14), (18), (21), Yields and interest rates from *Financial Statistics*, dividends and interest are net of income tax and profits tax savings on interest, average rate of increase of consumer price index is subtracted to derive real rates.

Lines (11), (15) calculated from data on yields and price changes in *Financial Statistics*, with income tax at standard rate deducted from yields.

external equity is not much different from the average real return, in net dividends and capital gains, from holding equity (about 5½ per cent) The costs of internal equity (line 4) are lower than those of external equity at the beginning and end of the period when there was discrimination against dividends, in 1966–9 it is as low as 4 per cent Line (5) shows that giant firms (assets of £500 million) had costs of internal equity about 1 per cent lower than small firms (£5 million) though this difference seemed to be diminishing Since a high pay-out ratio boosts valuation, firms with a pay-out ratio half as great again as the average (line 3) had substantially lower costs of capital (line 6) Earlier analysis of possible exaggeration in the Table I estimates of the effect of pay-out on valuation suggests that the reduction of the cost of capital for firms with a high pay-out ratio may be somewhat overstated The 'yield-form' model gave very similar estimates of the cost of equity capital The rather unsatisfactory TSLS tests gave slightly higher estimates Finally, the apparently rather higher estimates for the cost of external equity at the beginning of the period can partly be attributed to the lower pay-out ratio, as the calculations assuming a constant pay-out show (see line 8)

The estimated cost of debenture finance (line 13) is around 5 per cent, except in the period 1955–9, and this is about 1 per cent below the cost of equity But it is well above the real interest cost of debentures[1] (line 14) or the yield from holding debentures (line 15) so the 'gearing' effect on shareholders' profits is reflected in these estimates Even though the tests for non-linearity did not confirm increasing risk, in the final period the rather larger discrepancy between the estimated cost and the real interest cost may reflect concern about the substantial increase in gearing (line 12)

As a direct result of the high negative coefficients for preference stock in the valuation equation the estimated costs of preference capital were in general well above the costs of external equity the negligible use of preference finance (line 16) suggests that firms felt this finance to be expensive

The estimated cost of bank borrowing (line 20) fluctuates around 4 per cent, again way above the real interest cost of the borrowing (line 21) Even so, the effect of increasing risk on the cost of bank borrowing did not show up within the range of variation in the sample The TSLS coefficients suggested that these estimates of the cost of bank borrowing may in fact be rather low (1 per cent or so) in which case the cost of debentures and bank borrowing appear similar

Lastly the average cost of trade credit (line 23) is estimated to be very low, though the tests for non-linearity (see p 224) suggest that the marginal

[1] The nominal interest cost after deduction of both income tax and profits tax or corporation tax, and less the average rate of inflation in the period concerned

cost may be greater. Certainly the cost appears rather greater at the end of the period, when a lot of trade credit was being used (line 22)

The outstanding feature of the financing proportions is the drastic fall in the proportion of finance from internal sources in the late sixties (line 1); in turn reflecting the fall in over-all profitability (see line 27)[1]

The estimated average cost of capital (line 24) fell from over 5 per cent in the mid-fifties, to about 4½ per cent by the late sixties, generally remaining close to the average ratio of profits to total market valuation (line 26). This fall can be attributed to the reduced cost of internal equity, particularly under the corporation tax The average cost of capital fell much less than the sample measured rate of profit (line 27), and by the end of the period was probably greater than the true average rate of profit Presumably the prospect of valuation by the stock market of the profits from projects at less than the investments cost could reduce the incentive to accumulate capital Certainly any attempt to maintain the flow of internal finance by reducing dividends will substantially reduce stock market valuation, particularly when there is no tax discrimination against dividends. Similar adverse effects of raising external finance were not apparent over the period studied and it was not possible to identify rising marginal costs of borrowing for firms Nevertheless, the true costs of borrowing were well above the apparent interest costs and did not fall at the end of the period despite the greater tax concessions for debt interest

Corpus Christi College, Oxford

APPENDIX

(1) *List of companies in sample*

Where a bracketed year appears it refers to the *last* year the company appears

Food Bovril, Cerebos (1967); H P Sauce (1960), Mackintosh (1967), Rowntree, Spillers, Associated Biscuit Manufacturers, Reckitt and Colman.

Chemicals Albright and Wilson, Aspro-Nicholas, Boots, British Drug Houses (1966), British Glues (1967); British Paints (1963), Castrol (1965), Coalite and Chemical Products, Goodlass Wall & Lead Industries, Imperial Chemical Industries, International Paints (Holdings) (1967), Laporte Industries, Monsanto Chemicals (1967), Yardley (1965), Unilever

Non-electrical engineering· A P V Company, Averys, Babcock and Wilcox, Baker Perkins, British United Shoe Machinery, British Rollmakers, W H. Allen (1966), Glenfield and Kennedy (1964), Hopkinsons, Stone-Platt Industries, Ransomes, Sims & Jeffries, Renold Chain, Skefco Ball Bearing (1966), Newton Chambers, Gardner (L), John Thomson, G & J. Weir, Vickers

Electrical engineering Associated Electrical Industries (1966), British Insulated Callender's Cables, Ever Ready, Midland Electric, A. Reyrolle, C. A. Parsons (1967), Morgan Crucible, A. West.

[1] For more detail and discussion see Glyn and Sutcliffe (1972) chapters 3 and 6

(2) *Notes on the accounting data*

The accounting data processed by the Board of Trade has been thoroughly described in numerous publications, perhaps the most useful description being in Appendix C of Singh and Whittington. The following points on the construction of variables used here may be noted.

(i) Prior year adjustments are subtracted from Profit retained in Reserve in order to derive the figure for retentions best reflecting the year's activity Receipt of investment grants is not included (but see p 234)
(ii) Minority interest is not included in ASSETS, nor returns to such interests in DIST* or RET, since there is no way to split the returns between distributed and undistributed profits, or the capital between debt and equity.
(iii) Book value of equity includes provisions and future tax reserves on the grounds that in general neither represent an outsider's claim on the company's assets.
(iv) Current taxation is excluded from the company's ASSETS on the grounds that it varies between companies largely as a result of the timing of their accounting years
(v) Values for retentions and external capital raised prior to 1954 were derived from the Extel Cards for use in weighted averages for the early years.
(vi) Various errors and omissions in the data led to Reckitt and Colman and Baker Perkins being omitted in the first three years

(3) *Taxation adjustments*

To convert series for retentions and interest payments into 'dividend equivalence', such additional taxation must be deducted which would have been due had the retentions or interest been used to pay additional ordinary dividends The following elements of the company tax system are relevant for these calculations

(i) *Income Tax/Profits Tax 1951–64*

(*a*) Companies paid debenture and bank interest gross of income tax at the standard rate (t_i) and were not liable to any tax payments as a result of these interest payments
(*b*) All profits, after deduction of interest and depreciation allowances, were taxed at the standard rate of income tax plus the rate of profits tax on undistributed profits (t_u).
(*c*) Ordinary and preference dividends were liable to profits tax at the differential rate on distributed profits applied to their value gross of income tax (t_d is the *extra* rate of profits tax on distributed profits)

If O is ordinary and preference dividends net of income tax, E is gross earnings before interest and tax, but after depreciation, I is gross debenture and bank interest, and R is net retentions, then the tax system described above results in the following expression for O.

$$O = (E-I)(1-t_i-t_u)-R-Ot_d/(1-t_i)$$

$$O = [(E-I)(1-t_i-t_u)-R](1-t_i)/(1-t_i+t_d)$$

Thus retentions are multiplied by a factor $(1-t)/(1-t_i+t_d)$ to convert to dividend equivalence, and gross interest is multiplied by $(1-t_i-t_u)(1-t_i)/(1-t_i+t_d)$ So unit net interest payments yield a tax saving of (net interest—net dividend equivalent)

$$1-\frac{(1-t_i+t_u)}{1-t_i+t_d} = \frac{t_d+t_u}{1-t_i+t_d}$$

(ii) *Corporation Tax 1965 to 1968*

Treatment of debenture and bank interest was unchanged but

(*a*) All profits after deduction of interest and tax allowances, were taxed at the rate of corporation tax (t_c).

(*b*) Ordinary and preference dividends were liable to a withholding tax at the standard rate of income tax t_s on dividends paid after April 1966.

Using the same symbols as were used above:

$$O = (E-I)(1-t_c)(1-t_s)-R(1-t_s)$$

Thus retentions are multiplied by $(1-t_s)$ to reduce them to net dividend equivalence and gross interest is multiplied by $(1-t_s)(1-t_c)$. But *net* interest payment yields a tax saving of t_c in the form of higher total distributions (net interest plus dividends) as compared with the funds being used to pay extra dividends. Series for the tax factors reducing retentions to dividend equivalence, and for the tax savings on debt interest are given in Tables (*e*) and (*a*) respectively.

REFERENCES

DAVENPORT, MICHAEL, 'Leverage and the cost of capital', *Economica*, May 1971

MODIGLIANI, F. and MILLER, M. (MM (1958)) 'The cost of capital, corporation finance and the theory of investment', *American Economic Review*, June 1958

—— (MM (1961)) 'Dividend policy, growth and the valuation of shares' *Journal of Business*, Oct. 1961

—— (MM (1966)) 'Cost of capital to the electric utility industry', *American Economic Review*, June 1966.

GLYN, ANDREW and SUTCLIFFE, BOB, *British Capitalism, Workers and the Profits Squeeze*, Penguin, London, 1972

PARK, R. E., 'Estimation with heteroscedastic error terms', *Econometrica*, 1966

SINGH, A. and WHITTINGTON, G., *Growth, Profitability and Valuation*, Cambridge University Press, 1968

C.E.S. PRODUCTION FUNCTIONS IN BRITISH MANUFACTURING INDUSTRY: A CROSS-SECTION STUDY[1]

By TERENCE M. RYAN

Introduction

THIS article reports an exploratory attempt to estimate production functions for certain sectors of British manufacturing industry, the primary objective of the study being to determine the extent of the inter-industry *variation* in the elasticity of substitution between labour and capital.

The existence of such inter-sectoral differences in the substitution elasticity has several policy implications—for example—in providing a guide to the inter-industry pattern of future manpower requirements (contingent on projections of relative factor prices), and also in projecting the functional distribution of income.

The potential usefulness of estimates of the parametric coefficients of sectoral production functions, along with the consideration that the study of British production functions has tended to be a neglected field, together provided the motivation for the present study.

The paper is in two sections, the first of which reports on the indirect estimates of the substitution elasticities and compares them with similar estimates derived from other comparable studies The second section of the paper describes a procedure used to estimate the production functions directly, by means of a hill-climbing search algorithm The characteristics of the resulting estimates are then described.

The model: specifications and inputs

Following earlier studies in the field of empirical production function estimation—e g. Arrow *et al.* (1961), Hildebrand and Liu (1965), and Dhrymes (1965)—we start from the assumption that the production function displays constant elasticity of substitution between labour and capital. This is just about as general an assumption as we can make within the current theoretical framework, using only two factors of production.

Specifically, we examined seven major sectors of British manufacturing industry Engineering, Chemicals, Textiles, Electrical, Food, Building, and Printing. The sectors covered, though not exhaustive, were broadly representative of the range of manufacturing processes in the U.K

Observations were taken, at individual company level, in each of the

[1] Grateful acknowledgement is made to the Institute of Manpower Studies for financial support, and to my wife, June Ryan, for computer programming advice.

seven sectors for 1968 (268 companies in all), and for 1969 (337 companies), and these constituted the sample for the study.[1]

The production process in each of the sectors was described by a C.E.S. production function of the form ·

$$Q = \gamma_i\{\delta_i K^{-\rho_i} + (1-\delta_i)L^{-\rho_i}\}^{-1/\rho_i}, \tag{1}$$

the three parameters, γ_i, δ_i, and ρ_i being assumed to vary from one sector, to the next

The elasticity of substitution of each function is given by

$$\sigma_i = 1/(1+\rho_i) \quad (i = 1, \ldots, 7). \tag{2}$$

Two independent approaches were made to the problem of estimating the substitution elasticities, and are described in detail in the following sections. The first follows the traditional 'indirect' method developed by Arrow *et al* (1961), which calls for inputs of value added per man, and wages per man, while the second attempts direct estimation of the function by an iterative search algorithm, which uses hill-climbing techniques to find least squares estimates of the three parameters.

These parameters, γ, δ, and ρ, represent respectively the efficiency of the technology, the capital-intensity of the technology, and the ease of factor substitution In order that the production function be economically meaningful, it is necessary to impose certain *a priori* sign and magnitude restrictions on the parameters, as follows

$$\begin{aligned} &0 < \gamma < \infty, \\ &0 < \delta < 1, \\ &0 \leqslant \sigma \qquad (\text{i.e. } -1 \leqslant \rho \leqslant \infty, \text{ since } \sigma = 1/(1+\rho)) \end{aligned} \tag{3}$$

Indirect estimation of the elasticity of substitution

The C E S. production function raises considerable problems of statistical estimation, because of the non-linearity of its parameters, and its consequent insusceptibility to least squares estimation. Previous workers in this field, interested primarily in the size of the substitution elasticity rather than in estimating the entire function, have overcome this difficulty by resorting to the indirect method of estimation proposed by Arrow *et al*, who examined the relationship

$$\log(Q/L) = \alpha + \beta \log W, \tag{4}$$

where Q/L is value added per man, and W is wages per man. Under somewhat restrictive assumptions it may readily be shown that the relationship in the above equation implies the C E S. production function,[2] and furthermore that the parameter β may be interpreted as the elasticity of substitution.

[1] See Appendix for a detailed account of the sampling procedure.
[2] See Arrow *et al* (1961), pp 228 ff.

Cross-section estimates of the elasticity of substitution were obtained by this method for seven major sectors of the British economy in the years 1968 and 1969, and the results are presented in Table I.

TABLE I

	1968				1969			
	σ_i	SE	R^2	No.	σ_i	SE	R^2	No
Engineering	1 64**	0 14	0 68	63	1 43**	0 14	0 59	78
Chemical	1 29**	0 10	0 91	18	1 19**	0 10	0 87	26
Food	1 19*	0 09	0 86	28	1 13*	0 07	0 86	42
Textiles	1 07	0 07	0 83	45	1 12	0·10	0 75	45
Electrical	1 01	0 09	0 82	31	0 86**	0 01	0 99	39
Building	0 51**	0 02	0 96	52	0 68**	0 08	0 52	68
Printing	0 45	0·55	0 02	31	0 91	0 09	0 73	39

SE = Standard error; No. — Number of companies in sample.

(i) The fourteen estimates of σ_i all have the expected (*a priori*) sign, and are of a reasonable order of magnitude;

(ii) They are all very well defined statistically (with the sole exception of Printing, 1968), exceeding their standard errors of estimate many times (i e. all are statistically significant at the 99 per cent level),

(iii) The hypothesis that the estimates of σ_i displayed statistically significant differences was tested by pairwise comparison of the estimates (t-test) In the 1968 sample 11 of the 21 comparisons were statistically significant at the 95 per cent level, while in 1969 12 of the 21 comparisons were statistically significant Thus there is considerable support for the hypothesis that the sectors covered display significant difference in their elasticity of substitution;

(iv) The hypothesis that $\sigma_i = 1$ (implying a Cobb–Douglas production function) was tested and, as can be seen from Table I, was rejected for most sectors. One asterisk denotes that this hypothesis was rejected at the 95 per cent level, two asterisks that it was also rejected at the 99 per cent level

The above estimates provide substantial evidence to support the hypothesis that a significant inter-sectoral variation does exist in the elasticity of substitution, in British manufacturing industry

The estimates derived here beg comparison with similar estimates in the cross-section studies of Arrow *et al* (1961), Minasian (1961), Solow (1964), Hildebrand and Liu (1965), and Dhrymes (1965), all of which are based on either inter-regional (U S.A) data, or inter-country data.

In making such a comparison, however, one must be mindful of the fact that the industrial classification used in the present study may not

correspond exactly with the Two-Digit Manufacturing Industries classification used by the above authors.

On the other hand, the production functions estimated here are probably closer to being genuine engineering production functions, in that they are estimated at a more disaggregated level. Inter-regional or inter-country data present a greater probability of different technologies existing within the sample.

A comparison with the previous studies may readily be made from Table II. The most obvious difference in the estimates is the uniformly

TABLE II

Comparison of present estimates with other studies

	Present study		*Arrow et al.*	*Minasian*	*Solow*	*Lin and Hildebrand*	*Murata Arrow*		*Dhrymes*
	1968	*1969*	*1961*	*1961*	*1964*	*1965*	*1953/6*	*1957/9*	*1965*
Engineering	1.64	1.43	0.97	0.31	0.61	0.60	—	—	0.12 to 0.25
Chemicals	1.29	1.19	0.90	—	0.14	1.25	0.84	0.83	0.31 „ 1.03
Food	1.19	1.13	0.93	0.58	0.69	2.15	0.72	0.73	0.56 „ 0.97
Textiles	1.07	1.12	0.80	1.58	1.27	1.65	0.79	0.83	0.68 „ 1.03
Electrical	1.01	0.86	0.97	1.26	0.39	0.78	—	—	0.10 „ 0.62
Building	0.51	0.68	1.08	0.59	0.32	1.28	0.85	0.86	0.49 „ 0.89
Printing	0.45	0.91	1.21	—	1.02	—	0.84	0.93	0.68 „ 1.11

— = not available

higher values derived in the present study, most of which are significantly greater than unity, whereas the U.S. and inter-country figures are typically less than unity. This is somewhat surprising, especially in view of the greater disaggregation in the present sample.

Direct estimation of the elasticity of substitution

Although direct estimation of the C.E.S. production function is a cumbersome undertaking, it is not impossible. Consequently, in view of the restrictiveness of the assumptions underpinning the indirect estimation technique,[1] it was felt desirable to attempt direct estimation of the function.

It is evident from equation (1) that the substitution of arbitrary values for δ_i and ρ_i enables one to generate least-squares estimates of γ_i, the remaining parameter.[2]

It is, moreover, possible to arrive at a least squares equation for *any* set of data by trial and error, the only constraint being the computational complexity of the exercise involved. Bearing the above two considerations in mind, the approach adopted was as follows

[1] Especially the assumptions that both labour and product markets are competitive. See Arrow *et al.*, p. 228

[2] Using a forced zero intercept

(i) The fact that *a priori* considerations place bounds on the admissible values of ρ_i and δ_i means that the feasible region for the parameters is as illustrated in Fig. 1. The first stage in the estimation was therefore to make an exploratory test of this feasible region by generating least-squares estimates of γ_i (together with the resultant R^2) for various combinations of ρ_i and δ_i. Thus, corresponding to each point in the feasible region there

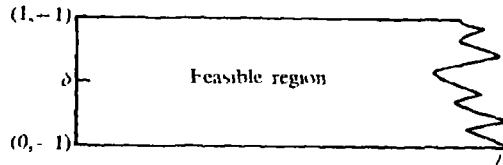



FIG. 1

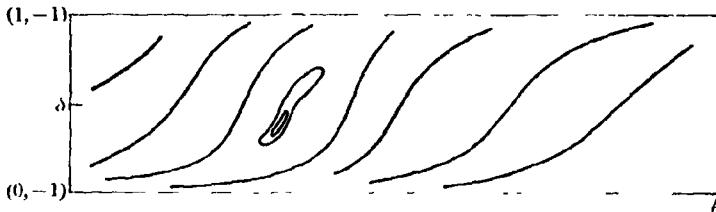



FIG. 2. Contours in R^2

exists a 'best' estimate of the underlying production function, as measured by R^2. What we are attempting to find is that combination of ρ_i and δ_i —and the resultant γ_i—which is the optimum optimorum.

(ii) The exploratory test of the feasible region consisted of taking 'soundings' in a grid pattern, and plotting the contours of the resulting R^2 values, as illustrated in Fig. 2. This confirmed R^2 values to be unimodal for all reasonable values of ρ_i and δ_i.

(iii) The final stage was to apply a hill-climbing search algorithm to the data in order to pinpoint the optimum optimorum more accurately. Here the Hooke and Jeeves Direct Search Method algorithm[1] was used, and the resulting estimates of the fourteen production functions, presented in Table III were obtained.

Measurement of the capital stock

The firm is regarded as a collection of capital goods, both tangible (machines, buildings, ...) and intangible (management skills, monopoly site, ...) which yields to its owners, the shareholders, a flow of income. The capitalization of this income stream was regarded as the most

[1] See Hooke and Jeeves (1961).

TABLE III

Interindustry production functions direct estimates

Industry / Capitalization Rate	Production function	R^2
Building (1968)		
12%	$Q = 1,116.8[0.0003K^{0.856}+0.9997L^{0.856}]^{1.168}$	$R^2 = 0.9709$
10%	$Q = 1,087.5[0.0006K^{0.759}+0.9994L^{0.759}]^{1.318}$	$R^2 = 0.9708$
8%	$Q = 1,100.2[0.0004K^{0.799}+0.9996L^{0.799}]^{1.252}$	$R^2 = 0.9708$
Building (1969)		
12%	$Q = 1,297.8[0.0002K^{0.899}+0.9998L^{0.899}]^{1.112}$	$R^2 = 0.9839$
10%	$Q = 1,292.9[0.0002K^{0.873}+0.9998L^{0.873}]^{1.145}$	$R^2 = 0.9838$
8%	$Q = 1,288.9[0.0002K^{0.853}+0.9998L^{0.853}]^{1.172}$	$R^2 = 0.9838$
Chemical (1968)		
12%	$Q = 622.7[0.0054K^{0.621}+0.9946L^{0.621}]^{1.610}$	$R^2 = 0.9923$
10%	$Q = 603.1[0.0056K^{0.600}+0.9944L^{0.600}]^{1.667}$	$R^2 = 0.9923$
8%	$Q = 626.5[0.0040K^{0.627}+0.9960L^{0.627}]^{1.595}$	$R^2 = 0.9923$
Chemical (1969)		
12%	$Q = 794.5[0.0026K^{0.695}+0.9974L^{0.695}]^{1.439}$	$R^2 = 0.9918$
10%	$Q = 786.7[0.0024K^{0.686}+0.9976L^{0.686}]^{1.458}$	$R^2 = 0.9918$
8%	$Q = 723.0[0.0038K^{0.617}+0.9962L^{0.617}]^{1.621}$	$R^2 = 0.9917$
Electrical (1968)		
12%	$Q = 780.9[0.0010K^{0.806}+0.9990L^{0.806}]^{1.241}$	$R^2 = 0.9776$
10%	$Q = 814.5[0.0004K^{0.901}+0.9996L^{0.901}]^{1.110}$	$R^2 = 0.9778$
8%	$Q = 796.0[0.0005K^{0.847}+0.9995L^{0.847}]^{1.181}$	$R^2 = 0.9777$
Electrical (1969)		
12%	$Q = 201.9[0.1810K^{0.108}+0.8190L^{0.108}]^{9.259}$	$R^2 = 0.9752$
10%	$Q = 193.9[0.1774K^{0.108}+0.8226L^{0.108}]^{9.259}$	$R^2 = 0.9752$
8%	$Q = 187.9[0.1746K^{0.108}+0.8254L^{0.108}]^{9.259}$	$R^2 = 0.9752$
Engineering (1968)		
12%	$Q = 862.4[0.0018K^{0.711}+0.9982L^{0.711}]^{1.406}$	$R^2 = 0.9932$
10%	$Q = 957.1[0.0005K^{0.830}+0.9995L^{0.830}]^{1.205}$	$R^2 = 0.9942$
8%	$Q = 958.0[0.0004K^{0.832}+0.9996L^{0.832}]^{1.205}$	$R^2 = 0.9942$
Engineering (1969)		
12%	$Q = 942.5[0.0020K^{0.690}+0.9980L^{0.690}]^{1.449}$	$R^2 = 0.9894$
10%	$Q = 890.7[0.0026K^{0.645}+0.9974L^{0.645}]^{1.550}$	$R^2 = 0.9892$
8%	$Q = 886.0[0.0024K^{0.641}+0.9976L^{0.641}]^{1.560}$	$R^2 = 0.9891$
Printing (1968)		
12%	$Q = 942.4[0.0018K^{0.706}+0.9982L^{0.706}]^{1.416}$	$R^2 = 0.9999$
10%	$Q = 922.2[0.0018K^{0.690}+0.9982L^{0.690}]^{1.449}$	$R^2 = 0.9999$
8%	$Q = 1,086.6[0.0003K^{0.842}+0.9997L^{0.842}]^{1.188}$	$R^2 = 0.9999$
Printing (1969)		
12%	$Q = 1,008.8[0.0052K^{0.505}+0.9948L^{0.505}]^{1.980}$	$R^2 = 0.9699$
10%	$Q = 1,022.8[0.0040K^{0.525}+0.9960L^{0.525}]^{1.905}$	$R^2 = 0.9699$
8%	$Q = 1,007.0[0.0043K^{0.505}+0.9957L^{0.505}]^{1.980}$	$R^2 = 0.9699$
Textiles (1968)		
12%	$Q = 699.2[0.0004K^{0.982}+0.9996L^{0.982}]^{1.018}$	$R^2 = 0.9820$
10%	$Q = 684.3[0.0005K^{0.927}+0.9995L^{0.927}]^{1.013}$	$R^2 = 0.9820$
8%	$Q = 695.9[0.0003K^{0.970}+0.9997L^{0.970}]^{1.031}$	$R^2 = 0.9820$

TABLE III (*cont.*)

Textiles (1969)		
12%	$Q = 751{\cdot}2[0{\cdot}0010K^{0{\cdot}801}+0{\cdot}9990L^{0{\cdot}801}]^{1{\cdot}248}$	$R^2 = 0{\cdot}9892$
10%	$Q = 742{\cdot}8[0{\cdot}0010K^{0{\cdot}778}+0{\cdot}9990L^{0{\cdot}778}]^{1{\cdot}265}$	$R^2 = 0{\cdot}9892$
8%	$Q = 760{\cdot}5[0{\cdot}0006K^{0{\cdot}829}+0{\cdot}9994L^{0{\cdot}829}]^{1{\cdot}206}$	$R^2 = 0{\cdot}9893$
Food (1968)		
12%	$Q = 424{\cdot}7[0{\cdot}0264K^{0{\cdot}389}+0{\cdot}9736L^{0{\cdot}389}]^{2{\cdot}571}$	$R^2 = 0{\cdot}9862$
10%	$Q = 418{\cdot}8[0{\cdot}0250K^{0{\cdot}385}+0{\cdot}9750L^{0{\cdot}385}]^{2{\cdot}597}$	$R^2 = 0{\cdot}9862$
8%	$Q = 420{\cdot}4[0{\cdot}0227K^{0{\cdot}388}+0{\cdot}9773L^{0{\cdot}388}]^{2{\cdot}577}$	$R^2 = 0{\cdot}9862$
Food (1969)		
12%	$Q = 651{\cdot}6[0{\cdot}0030K^{0{\cdot}676}+0{\cdot}9970L^{0{\cdot}676}]^{1{\cdot}479}$	$R^2 = 0{\cdot}9795$
10%	$Q = 652{\cdot}3[0{\cdot}0025K^{0{\cdot}678}+0{\cdot}9975L^{0{\cdot}678}]^{1{\cdot}475}$	$R^2 = 0{\cdot}9795$
8%	$Q = 650{\cdot}3[0{\cdot}0023K^{0{\cdot}675}+0{\cdot}9977L^{0{\cdot}675}]^{1{\cdot}481}$	$R^2 = 0{\cdot}9795$

appropriate measure of the capital embodied in the firm. This is essentially a Fisherian measure of capital input.[1] The choice of capitalization rate presented something of a problem, and in fact three alternative rates were used: 8 per cent, 10 per cent, and 12 per cent. As can be clearly seen from Table III, the estimates obtained were not significantly sensitive to the choice of capitalization rate. The estimates of δ, however, are not invariant to *units* of measurement of K and L. This accounts for the low values of δ presented in table III.

TABLE IV

Elasticity of substitution (direct estimates)

	1968	*1969*
Building	4·15	7·89
Chemicals	2·50	3·18
Electrical	1·01	1·12
Engineering	5·88	2·82
Printing	3·23	2·10
Textiles	13·78	4·50
Food	1·62	3·11

Corresponding to 10% capitalization rate.

The estimates of σ_i derived from the direct estimation of the production functions are presented in Table IV. Their most noticeable characteristic is that they all exceed unity. Furthermore, each direct estimate is larger than its corresponding indirect estimate.

[1] See Fisher (1930), pp. 14–15.

The estimates of γ, which measure the efficiency of the technology, should be treated with some reserve in this instance, because of the possibility of bias, introduced via the capital measure. This could arise in several ways; for example, to the extent that interest payments represent recent investment which has not, as yet, led to output increases, the estimates of γ derived would be biased downwards. Similarly a bias would result from a failure to take account of inventory changes over the course of the year; or from changes in the capitalization rate from one year to the next, brought about by changes in the shareholders' rate of time preference.

The exploratory nature of the present work did not warrant the considerable input of resources and time required to take account of such complications, especially as they do not bear directly on the elasticity of substitution.

Conclusion

To sum up, then: the over-all picture of British industry that emerges from the above work is that of a technology in which factor substitution is relatively easy, though significantly different from one sector to another. If the systematically higher estimates of σ obtained by direct estimation are to be given any weight, they would appear to imply a downward bias in the indirect estimates caused by a violation of the restrictive assumptions underpinning that method. This is an area which might usefully be explored further.

It is necessary, however, to emphasize the exploratory nature of the above work, and the need for refinement of the data and model specification before placing any heavy reliance on the estimates obtained.

University of Dublin, Trinity College

APPENDIX

The Data

Data source

One of the major difficulties in a study of this type is that of obtaining sufficiently disaggregated data on the manufacturing process. The present study is somewhat unusual in that it makes use of company reports as its primary data source. This calls for some justification and elucidation.

Company reports have traditionally been regarded with some suspicion by economists because of the anarchic nature of accounting practice. For example, the latitude allowed as to whether items go into the capital account or into the profit and loss account undermines the usefulness of the declared profit figure, while the conventional valuation of assets at cost makes an accurate assessment of the quantity of capital employed an impossible undertaking. Despite these and similar pitfalls, however, company reports can provide valuable information to the economist into the nature of the production process, provided he treads with care.

The sample

The sample consisted of a random selection of companies, 268 of them in the year 1968 and 337 in the year 1969, from those industrial manufacturing companies quoted on the London Stock Exchange. The breakdown between the seven industrial sectors can be seen from Table I. To a certain extent the categories, such as engineering, chemicals, etc, are fairly arbitrary. This, however, is due to the diversified activities of many large companies, which makes meaningful classification difficult.

Inputs

The indirect estimation of the production function called for inputs of output-per-man (Q/L) and wages-per-man (W)

Q/L The economist's definition of output—value added—does not correspond directly to any accounting concept. Value added, however, gets distributed as:

(i) payment to labour;
(ii) payment to capital, and
(iii) profit.

These three uses to which value added is put may readily be identified in company reports, being

(i) the aggregate remuneration of employees, plus national insurance contributions,
(ii) interest on borrowed moneys (both long-term loans plus bank overdrafts), and
(iii) total dividends paid plus transfer to revenue reserve.

The sum of (i), (ii), and (iii), divided by the number of employees (adjusted as described in the following paragraph) constitutes our estimate of Q/L.

W Company reports contain two items of relevance to this concept

(i) number of persons employed (weekly average); and
(ii) aggregate remuneration of employees for the full year.

Item (i) does not adequately measure labour input as it does not reflect changes in hours worked. Consequently such figures were adjusted for inter-sectoral differences in hours worked (40 hours being defined as an input of one man-week), and the resulting figures divided into item (ii) to yield an estimate of wages per man

REFERENCES

1. ARROW, K. J, CHENERY, H. B, MINHAS, B, and SOLOW, R. M, 'Capital-labor substitution and economic efficiency', *Rev Econ. and Stats* vol. **43**, pp 225–50.
2. BROWN, M., *The Theory and Empirical Analysis of Production* (NBER, 1967).
3. DHRYMES, P. J., 'Some extensions and tests for the CES class production functions', *Rev. Econ and Stats.*, Nov. 1965, pp. 357–66.
4. FISHER, I, *The Theory of Interest* (Macmillan, 1930).
5. HILDEBRAND, G. H., and LIU, T. C, *Manufacturing Production Functions in the United States* (Cornell, 1965).
6. HOOKE, R., and JEEVES, T. A., '"Direct search" solution of numerical and statistical problems', *Journal of the Association for Computing Machinery*, vol. **8**, 1961, pp. 212–29.
7. MINASIAN, J. R., 'Elasticities of substitution and constant-output demand curves for labor', *Journal of Political Economy*, June 1961, pp. 261–70.

8. MURATA, Y., and ARROW, K. J., Unpublished results of estimation of elasticities of substitution for two-digit industries from inter-country data for two periods, June 1965 (Reproduced in Brown (1967), p. 103).
9. SOLOW, R. M., 'Capital, labor, and income in manufacturing', in *The Behaviour of Income Shares* (Princeton, 1964), National Bureau of Economic Research.
10. TSURUMI, H., 'Non-linear two-stage least squares estimation of the CES production function applied to the Canadian manufacturing industries 1926–39, 46–67', *Rev. Econ. and Stats.*, May 1970, pp. 200–7.

CONSISTENT MEASURES OF IMPORT SUBSTITUTION

By GEORGE FANE

MANY recent studies have presented estimates of the contribution of import substitution to industrial growth [1] However, there is still no general agreement on the appropriate way to measure import substitution[2] and all the methods currently in use can lead to glaring inconsistencies when applied in situations involving aggregation over time periods or across industries. Fane [**10**] gave examples of the inconsistencies that have been obtained when aggregating over time periods and proposed a simple method which avoids these inconsistencies. Section I of this paper illustrates the inconsistencies which can occur when the conventional measures of import substitution are applied in situations involving aggregation across industries. Section II proposes a method for obtaining consistent results in this situation and Section III applies this new method to measure import substitution in manufacturing industries in Pakistan between 1954/5 and 1963/4, and compares the results with those originally obtained in a well-known study by Lewis and Soligo [**11**]

I. Defects of currently used measures of import substitution

Desai [**7**] has pointed out the need to distinguish between attempts to measure actual import substitution and optimum import substitution Optimum import substitution refers to the amount of import substitution that would have occurred had optimum policies been followed Several concepts of optimum import substitution are possible depending on whether one wishes to know what would have happened had optimum policies been followed throughout the period being studied, or just at the end of the period, or just at the start In common with the studies cited here, the present paper is mainly concerned with measuring actual import substitution using observed industry data on imports and domestic production If it were possible to calculate the growth paths of all relevant variables that would have been observed had optimum policies been followed, then it would obviously be possible to adapt *any* measure of actual import substitution to measure optimum import substitution

Most of the currently used measures of import substitution are adaptations of the procedure suggested by Chenery [**4**], according to which

[1] All the articles in the list of references, except [**3**] and [**15**], contain empirical estimates of import substitution. Nor is it claimed that this list is complete

[2] The measurement of import substitution has been discussed in [**3**], [**4**], [**5**], [**7**], [**9**], [**10**], [**11**], [**12**], [**13**], and [**15**].

import substitution in industry i is measured as:

$$(U_i^2 - U_i^1)Z_i^2. \tag{1}$$

Notation X_i domestic gross output in industry i

M_i imports competing with industry i

Z_i total supply (= demand) of output of industry i

Define:

$$U_i = X_i/Z_i. \tag{2}$$

There is an accounting identity.

$$Z_i = X_i + M_i. \tag{3}$$

From these definitions and identities one may derive the identity:

$$\Delta X_i = U_i^1 \Delta Z_i + (U_i^2 - U_i^1)Z_i^2, \tag{4}$$

where superscripts 1 and 2 denote the beginning and the end of the period being studied.

Lewis and Soligo used this identity in their study of structural change in Pakistan [11]. They used the first term on the right-hand side to measure the contribution of demand expansion to the growth of output and the second term to measure the contribution of import substitution In addition they separated the growth of total demand, ΔZ_i, into the growth of export demand and the growth of domestic intermediate and final demand.

Chenery used a different identity, but although Lewis and Soligo's procedure differs slightly from Chenery's their measure of import substitution is exactly the same Each of the terms on the right-hand side of the above identity is often expressed as a percentage of the total growth of output, ΔX_i

Lewis and Soligo used equation (4) to compute the sources of growth of output in twenty-six industries comprising all manufacturing industry in Pakistan for the sub-periods 1954/5 to 1959/60 and 1959/60 to 1963/4 and for the whole period 1954/5 to 1963/4 They were especially interested in total import substitution for each of three groups of industries those producing mainly consumption goods, intermediate goods, and investment goods. Their measure of import substitution for a group of industries is the sum of import substitution within each of the industries in the group, that is

$$\sum_i (U_{ij}^2 - U_{ij}^1)Z_{ij}^2,$$

where subscripts i and j now refer to industry i in group j.

Desai [7] pointed out that one might alternatively measure import substitution in group j by first aggregating imports and domestic production across industries in group j and then applying the Chenery/Lewis and Soligo measure; that is, measure import substitution for the group as

$$(U_j^2 - U_j^1)Z_j^2,$$

where
$$Z_j^t = \sum_i Z_{ij}^t \quad (t = 1, 2),$$
$$X_j^t = \sum_i X_{ij}^t \quad (t = 1, 2),$$
$$U_j^t = X_j^t/Z_j^t \quad (t = 1, 2)$$

Desai calls this procedure measure 2A and the Lewis and Soligo procedure measure 2B. Desai notes that the results will differ, that the ranking of different groups can be reversed by changing between the two measures; and that even more dramatic inconsistencies can sometimes occur for the Indian economy in the period 1951–63 Desai's results[1] appear to show that there was positive import substitution in each of the three groups producing consumption, intermediate, and investment goods. This was true whichever measure was used. However, according to measure 2A there was negative import substitution in all industries taken together. Did import substitution make a positive contribution to the growth of manufacturing industry in India in the period 1951–63? If one uses aggregated data the answer appears to be 'no', but if one uses disaggregated data the answer appears to be 'yes'.

Desai also considers two variants (denoted by 1a and 1b) of another method of measuring import substitution. According to variant 1a import substitution in group j is measured by [2]
$$(U_j^1 - U_j^2).$$
Variant 1b is
$$(U_j^1 - U_j^2)/(1 - U_j^1).$$

Negative values of these measures indicate positive import substitution Obviously both measures give the same sign, and if import substitution is positive according to either of these measures it must also be positive according to measure 2A
$$(U_j^2 - U_j^1)Z_j^2$$

Therefore, in the situations when measure 2A gives opposite signs for the contribution of import substitution to growth (depending on whether one uses aggregated or disaggregated data), variants 1a and 1b will also each give contradictory results This is illustrated by Desai's calculations according to either variant 1a or variant 1b there was positive import substitution in every sector of Indian industry in the period 1951–63, but negative import substitution for all sectors together [3]

In all the studies mentioned so far the expansion of domestic intermediate demand for the output of an industry is treated as a source of growth of the gross output of that industry. However, Morley and Smith [12] and [13]

[1] [7], p. 322, Table I, columns (4) and (5)

[2] Desai uses different notation and writes the definitions in a slightly different form See [7], p 318.

[3] See columns (1) and (2) of Table I, op cit

adopt a different procedure. they assume that the economy can be described by a Leontieff open model. Let A denote a matrix whose typical element is a_{ij}, the input of i per unit of gross output of j. Let r_{ij} be the typical element of the Leontieff inverse matrix $(I-A)^{-1}$. Any given vector of final demands $(F_1,.\ ,F_n)$ requires a total gross output (domestic and foreign) from industry i defined by:

$$Z_i^* = \sum_j r_{ij} F_j. \tag{5}$$

The domestic gross output is given by X_i, therefore the foreign gross output of industry i which is implicitly devoted to supplying the needs of the domestic economy is given by:

$$M_i^* = Z_i^* - X_i = \sum_j r_{ij} M_j. \tag{6}$$

Morley and Smith define U_i^* by·

$$U_i^* = X_i/Z_i^*. \tag{7}$$

Then·

$$dX_i = U_i^* \, dZ_i^* + Z_i^* \, dU_i^*. \tag{8}$$

Morley and Smith use the second term in this identity to measure import substitution in industry i.

Morley and Smith assert[1] that 'To replace an import, production must rise not only in the final processing industry, but also in the industries supplying its inputs and in their supplier industries, etc. Otherwise, there will be an induced rise in imported intermediates and/or a reduction in the supply of goods available for final demand in other sectors.

'In effect, the newly required intermediates were previously supplied indirectly or directly by the importation of the final product The replacement of implicit imports by domestic production is import substituting every bit as much as the direct substitution captured by [Chenery's measure], but will be missed by the usual definitions of imports and total supply.'

Against this one could argue that when, for example, M_1 is reduced and X_1 increased there is an increase in the intermediate demand for the outputs of the direct and indirect suppliers of industry 1. Suppose that industry j is one of these suppliers; the resulting increase in demand for the output of industry j represents a potential source of growth for j due to demand expansion every bit as much as an increase in final demand for the output of industry j, but will be missed by Morley and Smith's definition of demand expansion. Consider a hypothetical case where all final demands remain unchanged and where domestic gross output rises by one unit in industry 1 and remains constant in all other industries:

$$\Delta X_1 = 1,$$
$$\Delta X_j = 0 \quad (j = 2, 3, \ ).$$

[1] See [13], pp. 122–4

Imports must adjust to balance supply and demand for each industry:

$$\Delta M_1 = -1 + a_{11}$$

$$\Delta M_j = a_{j1} \qquad (j = 2, 3, \ldots)$$

Morley and Smith would record no import substitution and no demand expansion for industries 2, 3, .., etc. Chenery would record expansions in domestic intermediate demand for these industries, but that these sources of growth were exactly offset by negative import substitution imports rose by the full amount of the extra intermediate demands and domestic gross output failed to capture any of the potential growth.

It is arbitrary to say who is right; the present claim is only that Chenery's description is at least as informative as Morley and Smith's description.

Like the Chenery measure, the Morley and Smith measure of import substitution can yield inconsistent results in the sense that there could be positive import substitution in each industry and yet negative import substitution for all industries as a group.[1]

II. Consistent measures of import substitution for aggregation across industries

This section accepts Chenery's criterion for import substitution—that positive import substitution corresponds to an increase in the ratio of domestic gross output to total supply—and proposes a method for reconciling the different results which can be obtained using aggregated or disaggregated data

The proposal is that import substitution for industry i be measured in two parts· import substitution within the industry, denoted by S_i, and the extra contribution, S_i^*, of growth in industry i to import substitution in all industries. The total contribution of industry i to import substitution, S_i^T, is then defined using:

$$S_i^T = S_i + S_i^*. \tag{9}$$

Using formulas appropriate for small changes one may define dS_i and dS_i^* by:

$$dS_i = Z_i \, dU_i, \tag{10}$$

$$dS_i^* = (U_i - U) dZ_i, \tag{11}$$

where

$$X = \sum_i X_i,$$

$$Z = \sum_i Z_i,$$

and

$$U = X/Z.$$

S_i and S_i^* are obtained from dS_i and dS_i^* by integration.

[1] It is possible that $dU_i^* > 0$ for all i, and yet $dU^* < 0$, where $U^* = \sum_i X_i / \sum_i Z_i^*$.

The rationale for the definition of dS_i^* is that growth in an industry with a higher than average ratio of domestic production to total supply leads to an increase in this ratio for the whole group.

The contribution of import substitution to the growth of all industries, denoted by S, may be defined by applying equation (10) to aggregate data

$$dS = Z\,dU.$$

The advantage of the above definitions is that:

$$dS = \sum_i dS_i^T$$

since

$$\begin{aligned}\sum_i dS_i^T &= \sum_i Z_i\,dU_i + \sum_i (U_i - U)\,dZ_i \\ &= \sum_i (Z_i\,dU_i + U_i\,dZ_i) - U \sum_i dZ_i \\ &= \sum_i dX_i - U \sum_i dZ_i \\ &= dX - U\,dZ \\ &= Z\,dU.\end{aligned}$$

One may wish to consider two or more levels of aggregation over industries; examples have already been given of studies in which individual manufacturing industries have been aggregated into three sub-groups industries producing mainly consumer, intermediate, or investment goods These sub-groups can be aggregated to give totals for all manufacturing The approach set out in equations (9), (10), and (11) can be extended to yield consistent measures of import substitution in this more complex situation Let S_{ij} be import substitution within industry i in group j In small changes this may be measured by an expression exactly equivalent to the one in equation (10)

$$dS_{ij} = Z_{ij}\,dU_{ij}. \tag{12}$$

S_{ij}^* is the extra contribution of growth in this industry to import substitution in group j and is defined by

$$dS_{ij}^* = (U_{ij} - U_j)\,dZ_{ij} \tag{13}$$

This is analogous to equation (11).

S_{ij}^{**} is the extra contribution of growth in this industry to import substitution in all industries beyond its contribution to import substitution in group j.

$$dS_{ij}^{**} = (U_j - U)\,dZ_{ij}. \tag{14}$$

By analogy with (9) the total contribution of growth in industry i to import substitution for all industries is derived from

$$dS_{ij}^T = dS_{ij} + dS_{ij}^* + dS_{ij}^{**}. \tag{15}$$

The total contribution by group j to import substitution for all industries is given by:

$$dS_j^T = \sum_i dS_{ij}^T. \quad (16)$$

These formulas for measuring import substitution are consistent in the sense that:

(*a*) Import substitution for the aggregate of all industries is equal to the sum of the total contributions to import substitution in each individual industry:

$$dS = Z\,dU = \sum_j \sum_i dS_{ij}^T$$

(*b*) The total contribution by group j to import substitution for all industries (dS_j^T) is the same as the value that would have been obtained by treating group j as a single industry and using equations (9), (10), and (11)

(*c*) Import substitution within group j, defined as

$$dS_j = \sum_i (dS_{ij} + dS_{ij}^*)$$

is equal to the value that would have been obtained by treating group j as a single industry and using equation (10)

(*d*) The extra contribution of group j to import substitution by all industries defined by:

$$dS_j^* = \sum_i dS_{ij}^{**}$$

is equal to the value that would have been obtained by treating group j as a single industry and using equation (11).

All the measures proposed in this section have been defined for small changes The corresponding measures for finite changes can be obtained by integration This is illustrated by the example given in the next section.

III. Import substitution in Pakistan, 1954/5 to 1963/4

Lewis and Soligo collected and published[1] several very valuable data series for individual manufacturing industries in Pakistan in 1954/5, 1959/60, and 1963/4. The consistent measures of import substitution which were proposed in the previous section have been applied to Lewis and Soligo's data. This section presents and interprets the results and compares them with Lewis and Soligo's original estimates of import substitution.

One must make assumptions about the growth paths of imports and domestic production in each industry *between* the points for which one has observations in order to integrate the measures of import substitution for small changes which were proposed in Section II to obtain the corresponding measures for finite changes. The assumption made in the present

[1] See [11], Appendix A.

study was that domestic production and total supply in each i grew exponentially during each of the sub-periods 1954/5 to 1959 1959/60 to 1963/4. This is clearly an arbitrary assumption for e it implies that in general the growth of imports was not expc However, it appears to be no less plausible or more arbitrary tl alternative assumption

In order to evaluate the integrals involved in the various mea import substitution a computer programme was written which eff divided each sub-period into a further twenty 'micro-periods'. Tl of domestic production and total supply at the start and finish micro-period were derived from the assumption of exponential during the two sub-periods. However, within each micro-period was assumed to be linear. The reason for this procedure is t integrals can be evaluated analytically when growth is linear when growth is exponential. The justification is that the micro are so short that linear growth paths give a very close approxim exponential growth within each micro-period

The resulting measures of import substitution are presented in ((1) to (5) of Table I below. Column (6) gives Lewis and Soligo's estimates and column (7) gives the total increases in output as a s for comparison

In small changes columns (1) and (6) would be identical. In Tabl differ because of the different underlying assumptions about the paths followed. Nevertheless they give roughly the same pict growth in each sub-period within consumption industries and inv industries import substitution accounted for about a quarter to a all growth in the first sub-period, but was negative in the second. intermediate industries import substitution was positive in bo periods but was much more important in the second, when it ac for about a quarter of all growth, than in the first when it accou only about 5 to 8 per cent of growth The differences between colu and (6) are much more marked for the full period 1954/5 to 1963/ is not surprising since the larger the proportionate changes in any the greater will be the differences between the assumed underlying paths The column (1) estimates are clearly preferable for the ful since the assumed growth path[1] underlying the column (6) es violates the information on the levels of the relevant variables in

Column (3) gives roughly the same impression as column (1), e.

[1] When the column (1) estimates are expressed as a percentage of the growth they are identical to the estimates given in Fane [10], Table II, column (7), p pp. 4 to 6, explains the assumptions about the growth paths which are implicit and Soligo's estimates

the case of intermediate industries in the first sub-period. import substit tion within this group was much more than the sum of the contributio within each industry in the group The main reason for this was the fa growth of the jute textile industry, even in 1954/5 imports in this indust were negligible so that there were no opportunities for import substituti within the industry However, the growth of jute textiles substantial

TABLE I

Import substitution in Pakistan 1954/5 to 1963/4

(millions of rupees)

	$\sum_i S_{ij}$ (1)	$\sum_i S^*_{ij}$ (2)	$S_j = \sum_i (S_{ij} + S^*_{ij})$ (3)	$\sum_i S^{**}_{ij}$ S^F_j (4)	$S_j + \sum_i S^{**}_{ij}$ (5)	Lewis and Soligo (6)	Δ (
1954/5 to 1959/60							
Consumption	354	13	367	298	666	437	1,
Intermediate	35	92	127	- 102	26	54	[illegible]
Investment	86	10	96	367	−271	125	[illegible]
Total	475				420	616	2,
1959/60 to 1963/4							
Consumption	−102	- 24	127	531	404	−164	1,
Intermediate	192	−60	131	- 42	89	242	[illegible]
Investment	−55	34	−21	820	841	- 87	[illegible]
Total	34				347	−9	3
1954/5 to 1963/4							
Consumption	252	- 11	241	829	1,070	543	3,
Intermediate	226	32	258	−144	115	352	1,
Investment	31	45	75	1,187	-1 112	204	1
Total	509				73	1,099	6,

NOTE The basic data were taken from Lewis and Soligo (11), Appendix A Columns (6) and are taken from Lewis and Soligo, Appendix B The definitions and assumptions used in deriv columns (1) to (5) are explained in the text

reduced the ratio of imports to total supply for all intermediate industr taken together

Column (5) lists the total contribution of the industries in each gro to import substitution for the whole manufacturing sector These estimat differ dramatically from those given by Lewis and Soligo the column (estimates indicate that only just over 1 per cent of the total growth domestic manufacturing output in the full period 1954/5 to 1963/4 can attributed to import substitution, whereas Lewis and Soligo's estima (column (6)) of this contribution was 17·6 per cent

The estimates in columns (1) and (3) alone are also inadequate measu of the total contributions of each group of industries column (5) sho that the total contribution of all consumer goods industries in each per

greatly exceeded the amount of import substitution within these industries, while for investment goods the exact opposite was true. The total contribution of intermediate industries to import substitution for all manufacturing industries was less than the sum of import substitution within the group of all intermediate industries However, the difference is much less marked than for investment industries.

The explanation for these divergent results is simply that in 1954/5 domestic production was 77 per cent of total supply in consumer goods industries, 48 per cent of total supply in intermediate industries, and only 28 per cent of total supply in investment goods industries. Apart from increases in all three percentages in the first sub-period, the growth of consumer industries made an additional contribution to raising the percentage of total manufacturing supply which was met by domestic gross output, while the growth of the investment goods industries (and to a lesser extent that of intermediate goods industries) had an offsetting effect

In the second sub-period the percentage of domestic production in total supply fell for consumer and investment goods industries. That is, there was negative import substitution within these groups. Nevertheless the growth of consumer industries, in which domestic production now provided about 90 per cent of total supply, again tended to raise the percentage of domestic production in total supply for all manufacturing industries The corresponding proportion for investment goods industries in the second sub-period was still only about 35 per cent, so that the growth of these industries again tended to reduce the percentage of domestic production in total supply for all manufacturing industries.

The framework for measuring import substitution that has been proposed in this paper allows one to obtain consistent measures in situations involving the analysis of groups and sub-groups of industries and in situations involving several time periods The main contention of this paper is that these consistent measures are much more useful for analysing and describing structural changes than the existing measures, which can be inconsistent, paradoxical, and therefore very difficult to interpret

Harvard University and National Institute of Economic and Social Research, London

REFERENCES

1. BALASSA, BELA, 'Industrial development in an open economy. The case of Norway', *Oxford Economic Papers*, Nov. 1969.
2. BHAGWATI, JAGDISH N., and PADMA DESAI, *India Planning for Industrialization*, Oxford University Press, 1970.
3. BRUTON, HENRY J., 'The import-substitution strategy of economic development: a survey', *Pakistan Development Review*, Summer 1970.
4. CHENERY, HOLLIS B., 'Patterns of industrial growth', *American Economic Review*, Sept. 1960.

5. —— S. SHISHIDO, and T. WATANABE, 'The pattern of Japanese growth', *Econometrica*, Jan. 1962.
6. —— and L. TAYLOR, 'Development patterns among countries and over time', *Review of Economics and Statistics*, Nov. 1968.
7. DESAI, PADMA, 'Alternative measures of import substitution', *Oxford Economic Papers*, Nov. 1969.
8. —— 'Growth and structural change in the Indian manufacturing sector 1951–1963', *Indian Economic Journal*, Oct.–Dec. 1969.
9. EYSENBACH, M. L., 'A note on growth and structural change in Pakistan's manufacturing industry 1954–1964', *Pakistan Development Review*, Spring 1969.
10. FANE, GEORGE, 'Import substitution and export expansion', *Pakistan Development Review*, Spring 1971.
11. LEWIS, STEPHEN R., JR., and RONALD SOLIGO, 'Growth and structural change in Pakistan's manufacturing industry, 1954 to 1964', *Pakistan Development Review*, Spring 1965
12. MORLEY, SAMUEL A., and GORDON W. SMITH, 'On the measurement of import substitution', *American Economic Review*, Sept. 1970.
13. —— and GORDON W. SMITH, 'Import substitution and foreign investment in Brazil', *Oxford Economic Papers*, Mar. 1971.
14. STEEL, WILLIAM F., 'Import substitution and excess capacity in Ghana', *Oxford Economic Papers*, July 1972.
15. WINSTON, GORDON C., 'Notes on the concept of import substitution', *Pakistan Development Review*, Spring 1967.

DOES IT PAY TO TAKE A DEGREE?

The Profitability of Private Investment in University Education in Britain[1]

By ADRIAN ZIDERMAN

THE questions that are posed in this paper are: Has the large-scale post-Robbins expansion of higher education in this country conferred enhanced lifetime monetary earnings upon the large numbers of students that have undertaken higher education? If so, have these earnings been sufficiently high to offset any costs that students have incurred? In other words, what on average has been the rate of return on private investment in higher education?

Until recently, it has not been possible to speak positively about these issues since, apart from *ad hoc* salary data relating to particular professions or to non-representative samples of educated people, there has been lacking in this country nationally representative earnings data relating to various educational qualifications and levels. However, a special earnings survey in March 1968, carried out on behalf of the Department of Employment and Science by the General Register Office (now the Office of Population Censuses, and Surveys) and addressed to a sub-sample of educationally qualified persons enumerated in the 1966 Sample Census, has made available such information for Britain, for the first time. The data from this follow-up survey, which was specifically concerned with earnings for the financial year 1966/7, has been utilized to estimate private rates of return to university education in Britain, this paper presents these findings

The benefits of education

Do the educated gain positive earnings differentials over the less educated? A major finding of the Census follow-up survey[2] is that at each age earnings are higher the greater the level of educational attainment. This is shown in the charts, which give age-specific earnings data for the 1966/7

[1] An earlier paper, 'The rate of return on investment in higher education in England and Wales', *Economic Trends*, May 1971, written whilst the author was part-time Consultant at the Department of Education and Science, with the research association of Miss V Morris, concentrated on the economic return to society as a whole of investing in education, the focus of this paper is on the individuals themselves who obtain higher education As with the earlier paper, the methodology used and any views expressed must not be taken as representing those of the D E S The author gratefully acknowledges the research assistance of Mrs A Higgins, the early contribution of Dr I C R Byatt in initiating rates of return research at the D E S, and the helpful comments by Mr R J Allard, all the usual disclaimers apply.

[2] See *Survey of Earnings of Qualified Manpower in England and Wales 1966–67*, Statistics of Education, Special Series No 3, Department of Education and Science, H M S O., 1971, which contains details of sample, response, and specimen questionnaires

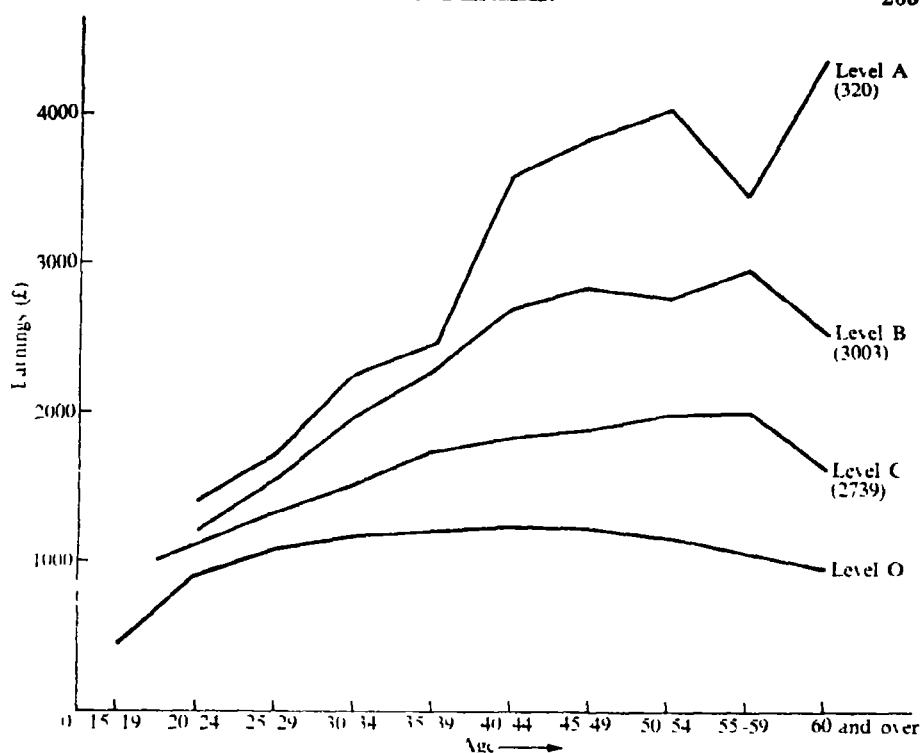



CHART 1

Average earning for levels of educational qualification, 1966/7 (Males)

KEY

Level A—Higher university degrees or equivalent

Level B—First degrees and all qualifications at this standard, including membership of certain professional institutions

Level C—Qualifications below first degrees, such as H N D, H N C, teaching certificates, and nursing awards

Level O—All males in working population.

Note Sample numbers shown in brackets

SOURCES

Levels A, B, and C 'Survey of earnings of qualified manpower in England and Wales, 1966–67', Department of Education and Science, 1971, Table 6

Level O Department of Health and Social Security (unpublished data)

financial year, for males and females (full and part-time), relating to three broad levels of educational qualifications [1] The charts also present age earnings profiles for employed persons generally, for comparison [2]

[1] The qualifications considered in the survey were those gained at age 18 or over and requiring study at a level above that required for G C E 'A' levels,

Level 'a'—higher university degrees or equivalent

Level 'b'—first degrees and all qualifications at this standard, including membership of certain professional institutions

Level 'c'—qualifications below first degrees, such as H.N D, H N C, teaching certificates, and nursing awards

[2] These latter earnings profiles inevitably include a small percentage of persons with the

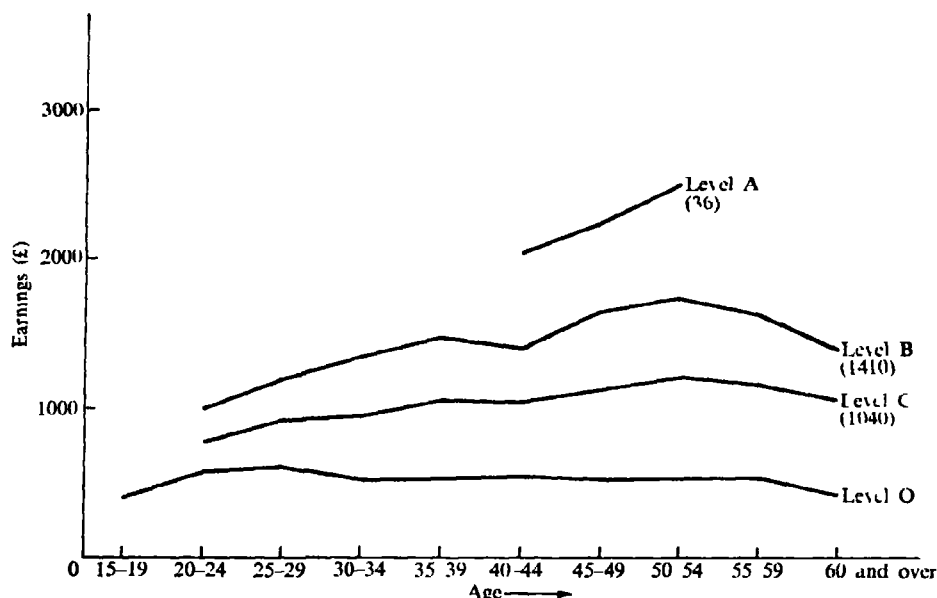



CHART 2

Average earning for levels of educational qualification, 1966/7 (Females)

KEY

Level A—Higher university degrees or equivalent

Level B—First degrees and all qualifications at this standard, including membership of certain professional institutions

Level C—Qualifications below first degrees, such as H.N D , H N C , teaching certificates, and nursing awards.

Level O—All females in working population.

Note Sample numbers shown in brackets.

SOURCES·

Levels A, B, and C 'Survey of earnings of qualified manpower in England and Wales, 1966–67', Department of Education and Science, 1971, Table 6.

Level O Department of Employment New Earnings Survey, 1968 (unpublished data)

Demonstrating that higher earnings are associated with extra education is not to suggest that these monetary benefits are the only ones accruing to individuals from education, or even necessarily the most highly prized. Clearly the benefits of education are many and multifarious. T. W. Schultz,

educational qualifications shown in the other profiles The age-earnings profiles for men generally were obtained from Department of Health and Social Security (D H S S) unpublished data and relate to a one-half per cent sample of persons registered under the National Insurance Act (The data refers to gross average annual 1966/7 earnings of males with at least one class 1 National Insurance contribution actually paid and at least 48 contributions credited.) This data source was not used for females because limitations in its coverage would have led to overestimated female average earnings. Instead, we used unpublished estimates of average weekly earnings for females in employment in G.B , from the Department of Employment's first New Earnings Survey, Sept 1968 The data were deflated on to a 1966/7 basis by use of the Department of Employment Index of Average Weekly Earnings of full-time women manual workers.

for example, has suggested the following three-fold classification of the benefits enjoyed by those who acquire extra education [1] education is an *investment* in higher future earnings, psychic income, and income in kind; there is a *consumer good* component satisfying consumer well-being in the present, and education is akin to a *durable consumer good* conferring future utilities over the lifetime of the educated

Not all economists would agree with this short list. Whilst all recognize that the benefits of education are not confined to the purely monetary ones, there still remains wide disagreement over what do constitute the benefits of education and how far the various benefits motivate individual choice. We do not enter into this debate here, but narrow our focus to investigating how far, if at all, the positive earnings differentials associated with university education have exceeded the costs incurred to the extent that students have anticipated a positive private rate of return on education, then we shall see whether these hopes have been fulfilled.

Data for calculating private rates of return

To estimate the rate of return on personal investment in a degree education, we systematically compare the monetary benefits of this education with the costs of acquiring it. Then, using conventional investment appraisal techniques, we compute an internal rate of return on this educational investment, which the individual may compare with going, or expected, rates of interest and with his time preference rate.

Consider an 18-year-old with good G.C.E. 'A' level attainments, who takes a job directly on leaving school. We may represent his expected lifetime earnings by the curve *OABC*, in Fig. 1, clearly, there are no earnings until age 18 and these then remain positive until assumed retirement at 69 Had he proceeded, instead, to university, then the curve *OADEGH* might depict his lifetime earnings expectancy; student maintenance grants and vacations earnings whilst studying (*ADE*), followed by very much higher earnings on graduation at age 21. Subtracting *OABC* from *OADEGH* to obtain the net monetary benefits of university education, since the area *ADEFCJ* is common, we are left with the positive area *FGHC* representing the positive earnings differential of those with university education and the negative area *DBFE*, the earnings forgone or opportunity cost of education. As university education in Britain is virtually

[1] T. W. Schultz, 'Investment in human capital *reply*', *American Economic Review*, Dec 1961. We are not here concerned with those additional (spillover) benefits of education, or externalities, not captured by the individual but accruing to others, these are imaginatively discussed by B Weisbrod in *External Benefits of Public Education*, 1964, and in 'Education and investment in human capital', *Journal of Political Economy*, Supplement Oct 1962, which also contains a classification of the personal benefits of education along different lines from that adduced by Schultz.

free at all levels, these opportunity costs are the only costs that the individual incurs. Hence we need only to estimate empirically the expected earnings at different educational levels, and obtain the costs by subtraction. It should be noted, in passing, however, that these opportunity costs are usually very high. For the educational qualifications gained full-time in Britain, earnings forgone per student are found to be of the same order of magnitude as overt (public sector) educational expenditures per student, a finding in line with the evidence from other countries.[1]

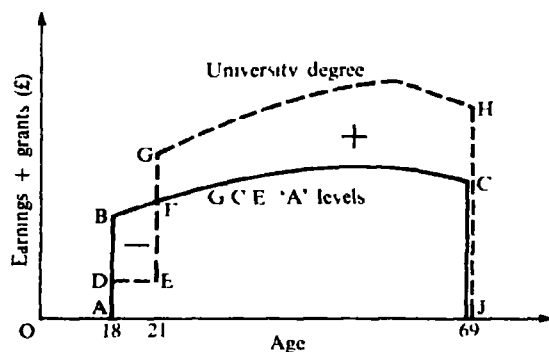



FIG. 1

We now consider in more detail the earnings data from which both benefits and costs estimates are derived. These data, as presented in the charts, are analysed in terms of three broad educational levels (corresponding to educational qualifications at the level of higher degrees, of first degrees, and at a level somewhat below first degrees). Since these groupings of qualifications are far too general and heterogeneous for our purpose of comparing private costs and benefits of university degree education, the data were sifted to retain only those with actual degree qualifications. Also excluded were persons possessing, in addition to degrees, certain professional qualifications not gained at universities, but requiring specified periods of on-the-job study and additional examinations (e g qualification in law, accounting, medicine), since it would not be possible to distinguish the separate effects on the earnings differentials of the degree and of the professional qualification. A further group to be excluded were those with occupations where significant non-monetary advantages exist (e.g. the military and clergy). Similarly, schoolteachers constituted strong candidates for exclusion, since it could be argued that their relatively low incomes in relation to persons with similar educational qualifications (as

[1] See, for example, the early study for the U S A. by T. W Schultz, 'Capital formation by education', *Journal of Political Economy*, Dec. 1960, and for India, V. N. Kothari, 'Factor cost of education in India', *Indian Economic Journal*, Apr.–June 1966.

confirmed in the Census follow-up survey) are offset by a large psychic income from their jobs. On balance, it was decided to include teachers within the analysis because their exclusion would seriously erode the number of usable observations; however, a sensitivity test was undertaken (and reported below) to see the effect of this decision on the magnitude of the rates of return estimates

With these exclusions made, the number of university degree observations remaining for use in the study were

	Males	*Females*
First degree	1,385	381
Master's degree	123	*
Doctorate	144	*

* Numbers of observations too few for analysis.

Utilizing the resulting average lifetime age-earnings profiles for degree levels, two types of calculations are presented in the sections that follow over-all rates of return for the whole education process from age 15 (average private rates of return) and rates of return to additional education (marginal private rates of return)

Average private rates of return

We compute average private rates of return for any given educational qualification by deducting from the expected lifetime average earnings stream for that qualification, the expected lifetime earnings profile of a 15-year-old school leaver (assuming compulsory schooling ends at age 15). In the absence of lifetime age-earnings information on those who leave school at age 15, we use as proxy data the mean earnings of men and women generally, shown respectively in Charts 1 and 2.

Since the individual will be interested in his disposable rather than pre-tax income, the latter is, fairly arbitrarily, adjusted by the application of representative tax rates.[1] The net-of-tax earnings streams are then corrected for the probability of labour-force non-participation, for unemployment and for mortality [2] One further correction to the earnings data

[1] Tax rates for the financial year 1966–7 were taken corresponding to the earnings data, the following family obligations for tax purposes were assumed, at various ages

Age 15–23	Single
24–5	Married, no children
26–9	Married, 1 child under 11
30–4	Married, 2 children under 11
35–44	Married, 2 children under age 11, 1 child 11–16
45–9	Married, 2 children over 16
50–69	Married, no children

[2] The after-tax age-earnings data were corrected for the *probability of non-participation* in the labour-force, using differential activity rates for educational qualifications at different levels obtained from the 1966 sample Census The earnings data derived from the follow-up

is necessary. Each age-specific earnings profile actually relates to a cross-section of different educated people at various ages, rather than to longitudinal data of the lifetime earnings of given individuals with particular educational attainments, which are simply unavailable. Since real earnings may be expected to rise over time, it is necessary to adjust these cross-sectional age-educational-earnings profiles to approximate to the lifetime earnings patterns of given educated individuals ageing over time. We conservatively assume that all earnings rise at the same annual rate of 2 per cent (thus widening absolute income differentials) though this will happen only if supply and demand for each type of educated manpower move generally in line.

With these adjustments made, the after-tax probability corrected earnings of a degree holder (Y_e) and of a 15-year-old school leaver (Y_o), for any year t, are

$$Y_{et} = [(1-T_t)y_{et}]u_e p_{et} s_{et}, \quad (1)$$

$$Y_{ot} = [(1-T_t)y_{ot}]u_o p_{ot} s_{ot}, \quad (2)$$

where y = pre-tax uncorrected earnings,
T = effective tax rate on earnings,
u = probability of employment,
p = probability of labour-force participation (see preceding footnote),
s = probability of survival,
and e and o relate respectively to degree education and school leaver aged 15.

The net of cost monetary benefits of degree education (B) for any year is obtained by subtracting (2) from (1), and adding average maintenance grants (G) and vacation and other earnings (V) whilst studying [1]

$$B_t = Y_{et} - Y_{ot} + G_t + V_t \quad (3)$$

Finally, the internal average rate of return on degree education (r) is found by solving for r in the following expression, which sets the present value of the net lifetime benefits stream of degree education equal to zero

survey were implicitly corrected for the *probability of unemployment* (and part-time employment) by the inclusion of actual earnings of all those who were economically active, since the economically active is defined in the Census to include the unemployed, the No Qualification profile for males contained a similar automatic correction, but that for females was corrected by the national average female unemployment rate. To allow for the *probability of survival* at each age, mean annual earnings were multiplied by a survival factor from the General Register Office life-expectancy tables; separate rates were obtained for the educationally qualified, derived from differences between survival rates for qualified and unqualified in the 1961 Census

[1] For undergraduates, the average maintenance grant per student in 1966/7 was £250, for graduate students a figure of £550 was taken, based upon the basic studentship grant of the major Research Councils Student earnings were assumed to accrue mainly from vacation work for undergraduates and from tutorial and other earnings for graduate students; an average annual figure of £80 was taken in each case

(assuming retirement is at age 69)

$$\sum_{t=15}^{69} B_t(1+r)^{15-t} = 0. \tag{4}$$

Rates of return derived in this way make the assumption that the whole of the earning differential associated with degree education is the result of that educational provision As we have noted above, to the extent that these earnings differentials are associated with a number of other causal factors, the procedure adopted would over-state the average rates of return results A large number of, mainly American, cohort and multi-variate[1] studies have attempted to deal with this issue, they have suggested alternative estimates of the fraction of the earnings differentials of the educated that are actually due to education, the most frequent ranging from 0·6 to 0 75 In the absence of any comparable estimates for Britain, we may make alternative assumptions regarding the probable size of the necessary correction, and test the sensitivity of our rates of return results to the assumptions made We report below the result of one such sensitivity test, assuming that 0·66 of the earnings differentials are due to degree education This implies that had the degree holder instead left school at age 15, his lifetime earnings pattern would have exceeded that of 15-year-old school leavers in general (because of his superior ability, family background, etc ,) by an amount equal to one-third of the earnings differential between degree holders and school leavers aged 15 In other words, we replace equation (2) above by the following

$$Y_{ot} = \{[y_{ot} + (1-\alpha)(y_{et} - y_{ot})](1-T_t)\}u_o p_{ot} s_{ot}, \tag{5}$$

where α is the fraction of earnings differentials *due* to education For the years of study when y_{et} is zero, y_{ot} has been raised arbitrarily to adjust for ability

Average private rates of return to degree education are presented in Table I, the main feature of which is the generally high level of the returns. No estimates are given for female higher degrees because of the paucity of observations in the survey It is noted that the 'ability' sensitivity test does not alter the general order of magnitude of the results, the sensitivity test for the exclusion of schoolteachers (not shown here) lowers all rates of return, but in no case by more than one percentage point The high return to females taking first degrees is to be explained not by high degree earnings but by the large degree *differentials* resulting from the particularly low average earnings of school leavers aged 15, partly the effect of lack of equal pay.

[1] The two best recent studies of this type attempting to adjust for non-education factors are G. Hanoch, 'An economic analysis of earnings and schooling', *Journal of Human Resources*, summer 1967, and D. C Rogers, 'Private rates of return to education in the United States a case study', *Yale Economic Essays*, spring 1969

The rates of return in Table I show the average return over the whole educational programme from age 15 that an individual, on average, receives; they are average both with respect to amounts of education and with respect to the cohorts in question. However, realistically, an individual will not have such a long-time horizon in view, rather, having achieved one qualification, he is likely to ask what on average will be the return from achieving an additional (higher) one These marginal (or incremental) rates of return are discussed in the following section

TABLE I

Average private rates of return on degree education from age 15, 1966–7

(per cent)*

	No 'ability' adjustment	*'Ability' adjusted*†
Males		
First degree	15 0	12 5
Master's degree	15 5	12 5
Doctorate	16 0	13 0
Females		
First degree	20 5	18 0

* All rates of return are rounded to nearest 0 5 per cent.

† Rates of return sensitivity tested for ability and other factors, by reducing earnings differentials by one-third.

Marginal private rates of return

These rates of return, average with respect to the cohort but marginal with respect to amounts of education, answer the question having completed a given level of education, what on average are the net monetary benefits from taking a little more? This somewhat more realistic profitability measure of private educational investment is calculated by solving i, in the following expression

$$\sum_{t=j}^{69} B_t(1+i)^{j-t} = 0, \qquad (6)$$

where B_t is now the net benefit of additional education and j is the age at which additional education begins.

Marginal rates of return are calculated here for first degrees (measured from G C.E. 'A' level) and for master's degrees and doctorates (both measured from first degree) Since age-specific earnings data relating to holders of G.C.E. 'A' levels were not obtainable by the Census follow-up survey, a less than satisfactory alternative estimate was used for males the salary scales of the Executive class of the Civil Service (for which

G.C.E. 'A' level forms a normal entry requirement), and assuming representative promotion patterns within the class. It might be argued that the use of the Executive class scales as a proxy for male 'A' level earnings would result in some overestimation, given the generally high level of Civil Service salaries; if this were the case, then the first degree marginal rates of return presented below would be somewhat *underestimated*. However, this data was thought to be an unsatisfactory proxy for female

TABLE II

Marginal private rates of return on male education, 1966–7

(per cent)*

	I *No 'ability' adjustment*	II *'Ability' adjusted*†	III *Adjusted for drop-out*
G.C.E. 'A' Level			
(from No Qualification)	10 0	8 5	Negative
First degree	22 5	20 0	16 5
(from G.C E 'A' Level)	(23 5)	(21 5)	(18 5)
Master's degree	20 0	16 5	Negative
(from first degree)	(19 0)	(16 0)	(Negative)
Doctorate	19 5	16 0	2 5
(from first degree)	(14 5)	(11 0)	(Negative)

* All rates of return rounded to nearest 0 5 per cent.

† Rates of return sensitivity tested for ability and other factors, by reducing earnings differentials by one-third.

NOTE · Figures in brackets are sensitivity estimates of the rates of return with schoolteachers *excluded*.

G C.E. 'A' level earnings generally since, given the existence of equal pay in the Civil Service and the rather lower level of occupations that females with 'A' levels are likely to hold, they would considerably overstate female 'A' level earnings. In the absence of a usable age-earnings profile for female G.C.E. 'A' level earnings and with the small number of female master's degree and doctorate observations, we present marginal rates of return for males only.

These are given in Table II, which shows rates of return for G C E. 'A' level as well as degrees. Referring first to columns I and II, we note the high rates of return shown for all degree qualifications; the 'ability' sensitivity test does not change the general order of magnitude of the returns. Separate rates of return are given in brackets, for schoolteachers excluded: these results might be relevant in the case where the student has already some occupation in mind, not involving schoolteaching. The effect, however, is minimal, except in the case of the doctorate—since there are few

teachers with doctorates, the exclusion of schoolteachers raises the first degree profile but leaves that of the doctorate profile unchanged, thus reducing earnings differentials.

Although the recorded rates of return for G.C.E. 'A' levels are lower, since they constitute an important stepping-stone to university education, and so carry a high option value to continue with education (as discussed by Weisbrod, but not estimated here),[1] they are somewhat underestimated.

Not all students successfully complete their courses; the final column of Table II presents the column I results adjusted for the probability of dropping out;[2] in the absence of data on the earnings of part-completers for Britain, we have assumed that the drop-out student gains no monetary advantage from his uncompleted course. Maglen and Layard, discussing the drop-out issue in connection with engineering courses,[3] distinguish between potential students who are 'optimistic' and those who are 'realistic'. The optimistic individual, contemplating extra education, may disregard the possibility of dropping out, whereas the realistic student may assume that he has only an average chance of completing his course However relevant, the effect of the drop-out adjustment is dramatic—only the first degree courses now stand out as good private investments.[4]

Some conclusions

What conclusions may we draw from the foregoing results? Investment in a first degree, under present free tuition and maintenance grant arrangements, offers high rates of return to the individual, and in higher degrees too, if the possibility of dropping out is not taken into account.

The high rates of return available from first degree education may suggest a considerable amount of private under-investment in university

[1] B. A. Weisbrod (1962), op. cit.

[2] The drop-out adjustment for first degrees is derived from *Enquiry into Student Progress*, U.G.C., 1968, and for post-graduates from E Rudd and S. Hatch, *Graduate Study and After*, 1969

[3] L. Maglen and R. Layard, 'How profitable is engineering education?', *Higher Education Review*, spring 1970, p 61.

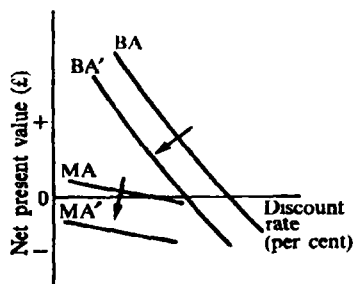



Fig 2

[4] It may seem strange that, given the rather similar results for all degrees in column I, the drop-out adjustment should have such drastic effects on higher degrees, but not on first degrees, this is due to the low net present values of higher degrees. In the case of first degrees, lifetime undiscounted net benefits are high (£27,000) but fall sharply as the rate of discount is raised the drop-out adjustment shifts the net present value curve to the left, as shown in the diagram, but affects the internal rate of return only mildly. The master's degree, however, gives low lifetime undiscounted net benefits (£5,500), which fall only slowly with raised discount rates; the drop-out adjustment moves the net present value curve downwards, but net present values are now negative for all discount rates, as shown

education. How can such high rates be explained? Although some have pointed to capital market imperfections, preventing private borrowing to finance education, as an explanation of these high private rates of return on first degrees, this is implausible for Britain, where fees are paid and generous maintenance grants are available for most students Indeed, it is not a problem of lack of demand for first degree places but rather of extremely high excess demands for these free university places, which are rationed by raising actual entry requirements (in terms of G C E 'A' level grades) far beyond those widely in use some years ago.

Capital market imperfections may be of more relevance for graduate education, where the number of studentships are limited We may explain the diverse results for higher degrees either by assuming graduate students are optimistic about their chances of successfully completing their courses or, if realistic about the possibility of drop-out, they put a high value on the psychic income available from those jobs to which a graduate degree qualification may lead It should be noted that the rates of return presented on master's degrees of all types may be a poor guide to the return available from the new highly vocational taught master's degrees offered by many universities in recent years

What light do these results shed on the vexed question of the full or partial replacement of maintenance grants by student loans? The reported high rates of return must be treated with some caution in this connection for two reasons Firstly they *include* grants, which count heavily in the total monetary benefits of education - since, received in the early years of the life-cycle, grants are affected little by the discounting process To test the feasibility of a loans scheme replacing grants it would be necessary to compute private rates of return *excluding* grants.

Secondly, since these rates of return are based on age-specific mean earnings by educational qualification at each age, there will be a spread of earnings around the mean value Given that most income distributions are skewed to the right, the majority of individuals in any age-education category will earn *less* than mean earnings and also their earnings differentials are likely to be less than the mean. We have been interested in private rates of return relating to the expected value of earnings at each age and have therefore used mean earnings, but with the result that the rates of return shown in Tables I and II, may not be achieved by the majority of individuals gaining those qualifications. The rates of return then will overstate the ability of the majority of individuals concerned to make repayments to a loans scheme.

We conclude on a cautionary note the private rates of return on degree education presented in this paper relate to the pattern of earnings differentials operative in 1966–7, and the assumption is made that this relative

pattern remains unchanged over time. This will not be the case unless supply and demand for educated individuals move more or less in step. The extent to which they have done so over the period 1966–71, and any corresponding changes in the rates of return, will emerge from a future analysis of the earnings follow-up to the 1971 Census, now in progress.

Queen Mary College, University of London

ADAM SMITH'S CONCEPT OF ALIENATION

By ROBERT LAMB

A NUMBER of recent critical works have touched on Adam Smith's description of the various deleterious effects of detail factory labour upon workers as an important predecessor to Karl Marx's concept of alienation[1] However, only E G West in his article 'The political economy of alienation, Karl Marx and Adam Smith', has attempted a comprehensive analysis of alienation in Smith's thought[2] West compares three aspects of alienation described by Marx—powerlessness, isolation, and self-estrangement—with the various comments by Smith concerning labourers[3] West concludes that Smith, unlike Marx, did not think labourers were alienated in terms of feeling powerlessness or isolation West acknowledges Smith's belief that factory workers are alienated by being self-estranged. But he does so only with the proviso that this conclusion is at variance with Smith's more general opinion that extensive division of labour cannot but improve men technically, socially, and morally I intend to show, in contrast to West, that Smith believed workers in some ways were self-estranged, powerless, and isolated By keeping Smith's integrated system of moral philosophy in mind and the historic conditions of labourers in the various industries he describes, I hope to provide a deeper analysis of Smith's views on alienation and the general effects of commerce and industry upon individuals than has been previously presented The most damaging effects of industry were expressed in what West calls Smith's

[1] Fritz Pappenheim, *The Alienation of Modern Man* (New York, 1959), pp 82–4 Jacob Viner, 'Guide to John Rae's *The Life of Adam Smith*' in John Rae *The Life of Adam Smith* (New York, 1965) especially p 35 Ronald L Meek, 'The Scottish contribution to Marxist sociology' in *Economics and Ideology and Other Essays Studies in the Development of Economic Thought* (London, 1967), pp 34–50 William F Campbell 'Adam Smith's theory of justice, prudence and benevolence', *American Economic Review*, vol 57, 1, no. 2 (May 1967), pp 571–7 Duncan Forbes's chapter in Douglas Young and others, *Edinburgh In the Age of Reason* (Edinburgh, 1967), especially p 47, and see George Elder Davis, ibid p 24

[2] E G West, 'The political economy of alienation, Karl Marx and Adam Smith', *Oxford Economic Papers*, vol 21, no 1 (Mar 1969), pp 1–23, hereafter cited as 'Alienation' This article of West's followed his earlier investigation, 'Adam Smith's two views of the division of labour', *Economica*, vol 31, no 121 (Feb 1964), pp 23–32, and Nathan Rosenberg's critique 'Adam Smith and the division of labour, two views or one', ibid, vol 32, no 126 (May 1965), pp 127–39

[3] Adam Smith does not use the term 'alienation' except in the eighteenth-century manner meaning 'sale' However, he does use the concept of human or personal alienation which Marx was to make famous. It is Smith's use of the concept not the term which is being investigated in this article In a current work in progress I am exploring certain inter-relations between these two

'alienation passage'.[1] Despite its length, this passage is so important that I feel it must be quoted in full

> In the progress of the division of labour, the employment of the far greater part of those who live by labour, that is, of the great body of the people, comes to be confined to a few very simple operations, frequently to one or two. But the understandings of the greater part of men are necessarily formed by their ordinary employments The man whose whole life is spent in performing a few simple operations, of which the effects too are, perhaps, always the same, or very nearly the same, has no occasion to exert his understanding, or to exercise his invention in finding out expedients for removing difficulties which never occur. He naturally loses, therefore, the habit of such exertion, and generally becomes as stupid and ignorant as it is possible for human creature to become The torpor of his mind renders him, not only incapable of relishing or bearing a part in any rational conversation, but of conceiving any generous, noble or tender sentiment, and consequently of forming any just judgment concerning many even of his ordinary duties of private life Of the great and extensive interests of his country he is altogether incapable of judging, and unless very particular pains have been taken to render him otherwise, he is equally incapable of defending his country in war The uniformity of his stationary life naturally corrupts the courage of his mind, and makes him regard with abhorrence the irregular, uncertain and adventurous life of a soldier. It corrupts even the activity of his body, and renders him incapable of exerting his strength with vigour and perseverance, in any other employment than that to which he has been bred. His dexterity at his own particular trade seems, in this manner, to be acquired at the expense of his intellectual, social and martial virtues. But in every improved and civilized society this is the state into which the labouring poor, that is, the great body of the people, must necessarily fall, unless government takes some pains to prevent it.[1]

Alienation is not a tiny phenomenon restricted to the few workers who were exploited in a few factories The above lengthy description of alienation among detailed factory workers is prefaced by Smith's statement that this alienation could cause 'the almost entire corruption and degeneracy of the great body of the people', and is concluded with the remark that this alienation 'must necessarily' take place among 'the great body of the people' in every improved and civilized society unless government especially through education steps in to remedy it Again, later on this page Smith says in all earlier societies 'Invention is kept alive and the mind is not suffered to fall into that drowsy stupidity, which in a civilized society, seems to benumb the understanding of almost all the inferior ranks of people '[1]

Smith's 'alienation passage' is by no means unique or unconnected with the rest of his thought concerning the effect of commercial-industrial society on most men Smith refers to these alienating effects of detail division of labour in his Glasgow *Lectures*[2] as early as 1763 and Nathan

[1] Adam Smith, *Wealth of Nations*, ed Edwin Cannan, 2 vols (London, 1960), hereafter cited as *WN*, ii, pp 302–3

[2] Adam Smith, *Lectures on Justice, Police Revenue and Arms*, ed Edwin Cannan (London 1896), hereafter cited as *Lectures*, pp 255–8

Rosenberg[1] and Ronald Meek[2] among others have indicated that this was probably a focal point of Smith's spoken lectures from 1751 or even 1749—fully twenty-seven years prior to the publication of the *Wealth of Nations* For example in Smith's *Lectures*, he states,

> There are some inconveniences, however, arising from a commercial spirit. The first we shall mention is that it confines the views of men Where the division of labour is brought to perfection, every man has only a simple operation to perform, to this his whole attention is confined, and few ideas pass in his mind but what have an immediate connection with it [3]

Again in the *Lectures* Smith contrasts the house carpenter's variety of tasks with

> . a cabinate maker, that particular kind of work employs all his thoughts, and as he had not an opportunity of comparing a number of objects, his views of things beyond his own trade are by no means so extensive as those of the former This must be much more the case when a person's whole attention is bestowed on the seventeenth part of a pin or the eightieth part of a button, so far divided are these manufactures. It is remarkable that in every commercial nation the low people are exceedingly stupid The Dutch vulgar are eminently so, and the English are more so than the Scotch The rule is general; in towns they are not so intelligent as in the country, nor in a rich country as in a poor one [3]

Although in his early *Lectures* alienation is introduced in a general discussion of the division of labour, the majority of Smith's later comments concerning alienation or self-estrangement in the *Wealth of Nations* are presented in his discussion of the State and education Smith may have thought that a full-scale discussion of alienation as caused by advanced specialization of tasks in his description of the benefits of the division of labour at the beginning of the *Wealth of Nations* might confuse his readers. It might make them fail to realize those many positive achievements modern society had made because of the division of labour

Alienation is analysed in detail in his chapter on the State specifically because Smith argued that it is only the State through education which can alter this mass degeneration of the workers (managers and workers cannot and will not do it for themselves) Nevertheless, as the two previous quotations from the early *Lectures* demonstrate, this alienating effect, or extreme 'confining of view' of the labourer working on the eightieth part of a button', had concerned Smith for over two decades[4] and this was by no

[1] Nathan Rosenberg, Adam Smith and the division of labour two views or one op cit, pp 127 39

[2] Ronald L Meek, Smith, Turgot and the four stages theory *The History of Political Economy*, vol 3, no 1 (Spring, 1971), pp 9 27 See especially pp 16 24 where on p 19 Meek cites Adam Smith in 1755 in a document publicly defending the originality of his theories in his lectures against counter claims These were treated of at a length in some lectures which I have still by me and which were written in the hand of a clerk who left my service six years ago'

[3] *Lectures*, p 255

[4] *WN* ii, pp. 305, 308, and as previously cited, 302–3, *Lectures* pp 255–8 And see Adam Smith, 'An early draft of the Wealth of Nations in W R Scott, *Adam Smith As Student and Professor* (Glasgow, 1937), hereafter cited as *Draft*, see *Draft* pp 327–8

means an afterthought or recently acquired idea but one of his earliest, and most consistent themes throughout his writings His early lectures state,

> These are the disadvantages of a commercial spirit. The minds of men are contracted and rendered incapable of elevation. Education is despised, or at least neglected, and heroic spirit is almost utterly extinguished. To remedy these defects would be an object worthy of serious attention [1]

This remains an excellent summary of his later discussion of alienation in the *Wealth of Nations* [2]

More important, in both early and late works it is the alienating effects of modern society on the vast majority of the population that Smith analyses Alienation is not suddenly restricted to a tiny problem of a fraction of factory employees Not in the 'alienation passage' but a few pages later Smith re-emphasizes that

> The same thing may be said of the gross ignorance and stupidity which, in a civilized society, seem so frequently to benumb the understandings of *all the inferior ranks of people* A man without the proper use of the intellectual faculties of a man, is, if possible, more contemptible than even a coward, and seems to be mutilated and deformed in a still more essential part of the character of human nature.[3]

This description of the mutilation and deformed essential character of human nature is as clear a description of self-estrangement or alienation as any Marx was to offer

In his article West concedes that this self-estrangement of the detail factory labourer might be the one form of alienation Smith may have intended to depict [4] And West is willing to admit that Smith considered this condition to result from overly extensive division of labour in factory production. But West then emphasizes Smith's self-contradictory approval and disapproval of the division of labour

> Thus in Book I he argues that workers become 'slothful' and 'lazy' without the division of labour; compare this with his claim in the alienation passage that they become 'stupid and ignorant' with it [5]

Although West notes these direct contradictions several times he does not explain them The explanation would appear to be that Smith's statements about the socializing effects of the division of labour are based on his abstract model of society built up from conjectured individual propensities originating in man's innate tendency to truck, barter, and exchange. On the other hand, his critique of the division of labour's effect on detail factory labourers is drawn from observations of the real social effects of such institutions upon individuals.[6] In a forthcoming article of

[1] *Lectures*, p 259 [2] *Lectures*, p 251 and *WN* ii, pp 302–3, as previously cited

[3] *WN* ii, p 308, my emphasis [4] West, *Alienation*, pp 9–11 [5] Ibid, p 11

[6] See Glen R Morrow, 'Adam Smith, moralist and philosopher', *Journal of Political Economy*, xxxv (1927), pp 321–42 This difference between Smith's conclusions reached from his abstract theories and, on the other hand, from his direct observations runs throughout his works

mine[1] I have examined the implications of the conflicts of these two methods employed by Smith: one leading to an abstract theory of capitalism; the second to socialist criticism of existing society.

E. G. West in his article has chosen primarily Smith's positive views of the effects of commerce from theoretical statements about how the market ought to operate,[2] not those passages where Smith is actually describing the commercial industrial market he lived in.[3] For example, West argues from Smith's general theoretical view of the commercial model as this 'happiest and most comfortable . . . cheerful and hardy state to all the different orders of society'.[4] But this theoretical appraisal must be contrasted with Smith's actual description of existing 'conditions so justly lamented'. Smith occasionally uses 'misery', and 'want' to describe the feelings of the poor in their present conditions, but 'happy' and 'cheerful' he reserves only for their periods of amusement when they are not working.[5]

Although it would be inaccurate to omit his descriptions of cheerful class relations in his theoretical model of commercial society, for the study of his concept of alienation the evidence is contained almost exclusively in descriptions of the actual conditions of labourers, not in his abstract economic theory of relations in a hypothetical, progressive state. Smith's descriptions of alienation or self-estrangement of the detail factory worker stemmed from his own observations among the pin makers, loom, die, chemical, and munitions workers.[6] His alienation passages rest upon those observations from which he drew his recognition of class conflict, and his exploitation theory of wages,[7] profit,[8] and rent.[9] In contrast, his abstract economic theory assumed a theoretical harmony of class interests, and it was from this he developed his supply and demand theory of wages,[10] profit,[11] and rent,[12] which denied that exploitation was taking place and appeared not to consider the alienating effects of industry on workers at all.

[1] Robert Lamb, 'Adam Smith's system: sympathy not self-interest', *Journal of the History of Ideas*, 1973.

[2] West, 'Alienation', pp. 8–14. West repeatedly refers to Smith's first chapter of the *Wealth of Nations*, I, pp. 12–17, where this theoretical view of how the market ought to operate is most predominant.

[3] *WN* I, pp. 2, 54–5, 68–70, 74, 162. *WN* II, p. 308. And see especially *WN* I, p. 519. 'Commerce, which ought naturally to be, among nations, as among individuals, a bond of union and friendship has become the most fertile source of discord and animosity.'

[4] West, 'Alienation', p. 9.

[5] *WN* II, p. 318, 'the gaiety of public diversions'.

[6] *WN* II, pp. 302–3, previously cited. *WN* I, pp. 114, 138, 88, 74–6, 18–19.

[7] *WN* I, p. 74, and see Joseph Schumpeter, *History of Economic Analysis*, pp. 111, 189.

[8] *WN* I, pp. 74, 68–70, and Schumpeter, op. cit., pp. 190–1, 268, 331–4.

[9] *WN* I, p. 162, and Schumpeter, op. cit., pp. 190–1, 264–5, 268.

[10] *WN* I, pp. 77, 83, 111, and see Schumpeter, op. cit., pp. 111, 189–90, 268–70.

[11] *WN* I, pp. 62–3, 66, and see Schumpeter, op. cit., pp. 190–1, 268, 331–4.

[12] *WN* I, pp. 62–3, and see p. 164. And see Schumpeter, pp. 190–1, 264–5, 268.

While it may not be possible fully to reconcile Smith's conflicting positive and normative economic theories, his discussion of alienation provides a key for reconciling some contradictions in his views on the negative and positive effects of the division of labour. This reconciliation Smith accomplished himself by his historical continuum of progress in stages of social development[1] By this historical scheme the positive effects of division of labour in theoretical or actual primitive society (in the famous hypothetical deer–beaver exchange in the first division of labour) gave way only eventually to the detrimental alienating effects of extreme division of labour among modern factory workers[2] These same stages of progress are also suggested by Smith when he says alienation will further increase as the division of labour in factories increases[3]

Another reason for believing that Smith's historic stages concept explains what E G. West considers to be Smith's self-contradictions on alienation is that it directly corresponds to the explanation by Smith's two close associates, Adam Ferguson and John Millar, of the contrary effects of division of labour Their descriptions of the harmful effects of the extreme divisions of labour in alienating men, and the positive effects of the general divisions of labour in improving men, are very similar to Smith's. Ferguson frequently contrasts these two effects directly whereas Smith only contrasted them by implication Ferguson said, for example:

> The separation of professions, while it seems to promise improvement of skill, and is actually the cause why the productions of every art become more perfect as commercial society advances, yet in its termination, and ultimate effects, serves, in some measure, to break the bands of society, to substitute mere forms and rules of art in place of ingenuity, and to withdraw individuals from the common sense of occupation, on which the sentiments of the heart, and the mind, are most happily employed.[4]

Also in a similar fashion to Smith, Ferguson warns that extreme division of labour dismembers the human character,[5] while directly contrasting this with the great productivity which such division makes possible[6]

John Millar dwelt in greater depth than Smith or Ferguson on the psychological effects of extreme division of labour Millar concluded, 'There are limits beyond which it is impossible to push the real improvements arising from wealth and opulence '[7] Nor was this a passing interest of Millar's, for his biographer insists that his lectures over many years examined 'at some length the question of whether the progress of civiliza-

[1] See Smith's *Lectures*, pp. 14–15, 108–12, 161, *WN* ii, p 231.

[2] *WN* ii, pp 302–3, previously cited

[3] *WN* ii, pp 300–8

[4] Adam Ferguson, *An Essay on the History of Civil Society*, ed Duncan Forbes (Edinburgh, 1966), hereafter cited as *History*, p 181, and see generally part iv, section 1, and part v, section 3, 4

[5] Ferguson, *History*, pp 280, 353

[6] Ibid , pp 353

[7] Duncan Forbes, 'Scientific Whiggism, Adam Smith and John Millar', *Cambridge Journal*, vol. vii, no ii (Aug 1954), pp 643–50

tion can be continued without end, or whether it is subjected to certain limitations, from the nature of human affairs' [1] In addition to considering the alienating effects of intense divisions of labour, Millar and the Scottish School examined the more general social alienation resulting from commercial competition Smith said, 'Commerce, which ought naturally to be among nations, as among any individuals, a bond of union and friendship has become the most fertile source of discord and animosity '[2] Millar concluded that prosperous competitive conditions of modern commerce

> Contract the heart and set mankind at variance In proportion as every man is attentive to his own advancement, he is vexed and tormented by every obstacle to his prosperity and prompted to regard his competitors with envy, resentment and other malignant passions [3]

This more general alienation or self-estrangement theme is found earlier than Smith's *Lectures* in Rousseau's *Discourse on the Origin of Inequality* However, Smith directly attributed this alienation or self-estrangement to the process of the detailed division of labour which Rousseau's *Discourse* did not mention.[4]

We turn from Smith's concept of alienation, meaning the self-estrangement of workers (upon which even E G West is willing to agree Smith held strong views), to the question of whether Smith believed workers were alienated by token of their powerlessness and isolation On this question there may be some room for confusion since labourers could be powerless because of the existing inheritance system of entail and primogeniture, floods, droughts, etc , factors which need not result from advanced division of labour or modern commerce Similarly, workers could be isolated because they lost their traditional social cohesion when they were uprooted from rural districts and thrown into cities instead of being isolated exclusively or primarily by modern commercial divisions of labour. However, West, instead of considering these sorts of provisos for evaluating Smith's views of Isolation and Powerlessness as forms of alienation, asserts that Smith never believed workers were at all powerless,[5] and says, 'There is no evidence to show that Smith believed the detail workers felt isolated '[6]

In his effort to compartmentalize the various aspects of alienation into 'powerlessness', 'isolation', and 'self-estrangement', West has lost sight of their interconnection in the total system of Smith's moral theory His

[1] John Craig, *Life of John Millar* appended to John Millar, *Origin of the Distinction of Ranks* (Glasgow, 1806), 3rd edn , pp xviii–xix

[2] *WN* i, p 519, and *Lectures*, pp 259, 255

[3] Millar, p 387

[4] See Smith's comments on Rousseau's *Discourse on the Origin of Inequality* in letter to the *Edinburgh Review*, 1755–6 And see Duncan Forbes's Introduction to Ferguson's *History*, pp xxxi–xxxiii, on Rousseau and the Scottish School

[5] West, 'Alienation', pp 7–10

[6] Ibid , p 9

system of moral sentiments dictated that men gained their social values from society and if they became self-estranged (as West is willing to concede), then according to Smith's interconnected moral system, they become automatically isolated from others. West in a number of passages alluding to the connection between Smith's moral and economic writings appears to have misunderstood their interrelation entirely. For example, West claims that 'the whole process (of detail factory labour) was thus a coherent, positive, and constructive social process',[1] whereas Smith, as I have already quoted, says that detail factory labour 'renders him (a man) not only incapable of relishing or bearing a part in any rational conversation, but of conceiving any generous, noble or tender sentiment, and consequently of forming any just judgment concerning many even of the ordinary duties of private life'.[2] As labour became more divided into the most confined 'few simple operations, frequently but one or two', Smith nowhere argues as West does[3] that 'such communion provided men with the impartial spectators which they needed as a mirror of their actions'. On the contrary,[4] such activities completely destroyed their ability to sympathize with or to hearken to their internal impartial spectator.[5] Men became isolated by the very fact of their self-estrangement.[4] For in Smith's moral theory, one is only social to the extent that one is capable of sympathy.[6] Destroy his faculties for sympathy and the detail worker is by Smith's own definition isolated. Smith directly agreed with his colleagues Adam Ferguson and John Millar in this argument.[7]

The roots of this particular form of alienation by isolation are evident in Smith's ethical system of his *Theory of Moral Sentiments* in which individuals all confront each other as spectators.[8] Although indviduals have the ability to sympathize with others on the basis of viewing the external behaviour of others, there is no direct relation between men's internal selves.[9]

Smith believed that men adopt the role of 'impartial spectator' to judge the conduct of other men. Eventually, they form just judgements of themselves and their own behaviour through gradually applying this same impartial spectator's judgements they have used on others, finally on themselves.[9] Only from this external position do we come to know and feel with others, or to judge or feel happy with ourselves.[9]

On the very first page of the *Theory of Moral Sentiments* Smith lays

[1] West, 'Alienation', p. 10. [2] *WN* ii, p. 303, previously cited.

[3] West, 'Alienation', p. 10. And West is joined in this misunderstanding by Milton M. Meyers, 'Division of labor as a principle of social cohesion', *Canadian Journal of Economics and Political Science*, xxxiii, no. 3 (Aug. 1967), p. 432.

[4] *Lectures*, p. 257. [5] *WN* ii, p. 303, previously cited.

[6] Adam Smith, *Theory of Moral Sentiments* (London, 1907), p. 1.

[7] Ferguson, *History*, p. 218; Millar, pp. 380–2. [8] *TMS*, pp. 161–4.

[9] Ibid., pp. 162–4.

down the primary assumption that, 'we have no immediate experience of what other men feel'.[1]

> Though our brother is upon the rack, as long as we ourselves are at ease, our senses will never inform us of what he suffers. They never did, and never can, carry us beyond our own person, and it is by the imagination only that we can form any conception of what are his sensations. Neither can that faculty help us to this any other way, than by representing to us what would be our own if we were in his case It is the impressions of our own senses only, not those of his which our imagination copy.[2]

However, Smith's famous statements in the *Wealth of Nations* and in his *Lectures* make it absolutely clear that the vast majority of the public were, because of the detrimental effects of advanced division of labour, steadily losing their 'capacity to enter into the ordinary sentiments' of their fellows, losing the use of their physical, moral, and intellectual faculties This implied that men were becoming estranged and isolated from one another.[3] Even in his early *Lectures* Smith stated that 'In all commercial countries the division of labour is infinite and everyone's thoughts are employed about one particular thing . . . each of them is in a great measure unacquainted with the business of his neighbour '[4]

Having shown that Smith described alienation in terms of isolation and self-estrangement we must now consider whether he also conceived alienation in terms of powerlessness According to his descriptions the worker is effectively powerless: in disputes with employers, in determining the number of pieces in piece work, and the speed or length of time of his work in hourly labour, and powerless to decide the form of payment he will be given in credit or truck, cash or vouchers from private stores [5]

> The poor labourer who has the soil and the seasons to struggle with, and, who while he affords the materials for supplying the luxury of all the other members of the commonwealth, and bears, as it were, upon his shoulders the whole fabric of human society, seems himself to be pressed down below ground by the weight and to be buried out of sight in the lowest foundation of the building [6]

Smith describes such a worker as 'the lowest and most despised member of civilized society',[7] which contradicts West's assertion that 'There is no

[1] Smith, *Theory of Moral Sentiments*, pp 162–4

[2] Ibid, pp 3 4

[3] *WN* ii, pp 302–3, previously cited, *Lectures*, pp 255–9 *Theory of Moral Sentiments*, pp 3–4

[4] *Lectures*, p 257 See also *WN* i, p 519

[5] *WN* i, pp 74–6 Each of Smith's descriptions of workers' usual powerlessness is fully documented by economic histories of his time Henry Hamilton, *Economic History of Scotland in the 18th Century* (Oxford, 1963), pp 52, 346, 375–85 Thomas Johnston, *History of the Working Classes in Scotland* (Glasgow, 1921) chapter on Forced labourers, pp 72–84. T S Ashton, *An Economic History of England in the 18th Century* (London, 1966), see pp 224–35 See also Pinmakers document in *Tracts Relating to Trade*, Manuscript 13, volume of papers filed at the British Museum, p 816

[6] *Draft*, p 238 *Lectures*, p. 163. *WN* ii, p 308 *WN* i, pp 54–5.

[7] *Draft*, p 328 *WN* ii, p 308 Smith says they are called 'contemptible'

complaint in Smith that such subordination is undignified.'[1] West, therefore, appears to have inadequate justification for his statement that 'If one can speak of any subordination in Smith's economic system it is not to any one social group but to the consumer.' Smith distinctly describes these existing conditions as 'exhibiting so much oppressive inequality',[2] and he provided a detailed diagram of subordination.[3] Yet Smith usually approaches the subject of workers' powerlessness in disputes, contracts, etc., from the standpoint of descriptions of exploitation.

> In a society of a hundred thousand families, there will be perhaps one hundred who don't labour at all, and who yet, either by violence or by the orderly opposition of the law, employ a greater part of the labour of the society than any other ten thousand in it. The division of what remains too after this enormous defalcation, is by no means made in proportion to the labour of each individual. On the contrary those who labour most get least.[4]

In his *Wealth of Nations* he says, 'For every rich man there must be at least five hundred poor', and frequently he describes numerous aspects of the exploitation by merchants, manufacturers, landlords, and others of the labouring public.

E. G. West does not present the 'abundant evidence' that 'Smith believed labour commanded capital no less than vice versa'.[5] While this is partially true in the abstract, Smith's famous description was that

> It is not, however, difficult to foresee which of the two parties must, upon all ordinary occasions, have the advantages in the dispute, and force the other into a compliance with their terms. The master . . .[6]

In the long run the workman may be as necessary to his master as his master is to him, 'but the necessity is not so immediate . . .'[7] —'the necessity which the greater part of the workmen are under of submitting for the sake of present subsistence . . .'[8] And this dependence of workers must be seen in the light of Smith's famous statement that 'Nothing tends so much to corrupt mankind as dependency.'[9]

West's essential reason for not believing Smith could have considered labourers to be alienated by being powerless was 'Smith's general emphasis upon the private property in labour skill'.[10] West does not appear to realize that according to Smith, although workers theoretically possessed 'the most sacred property'[11] in their labour-skill, they were repeatedly referred to by Smith as actually: 'without property', 'the propertyless', or having 'none at all'.[12] Clearly the distinction between the men who had

[1] West, 'Alienation', p. 8.
[2] *Draft*, p. 328. And for explicit descriptions of oppression in the *Wealth of Nations* see *WN* i, p. 278, ii, p. 483.
[3] *Draft*, pp. 327–8. And see *WN* i, p. 2. *Lectures*, p. 163.
[4] *Draft*, p. 327.
[5] West, 'Alienation', p. 8.
[6] *WN* i, p. 74.
[7] *WN* i, p. 75.
[8] *WN* i, p. 76.
[9] *Lectures*, p. 155.
[10] West, 'Alienation', p. 9.
[11] *WN* i, p. 136 and p. 72, chapter 8, first sentence.
[12] *WN* i, p. 75.

'some property' and the vast majority who had 'none at all' did not refer to those who could not labour [1]

Smith's criticism of exploitation did not lead him to the Marxian conclusion that different classes or their functions should be abolished. However, Marx[2] relied upon Smith's descriptions of the effects of detail factory labour upon the workers and Smith's more general analysis of the wage–labour relation to capital owners for his entire concept of alienation [3] But it may not be generally realized that, in his view of labour as an activity 'in which man alienates himself, . . a labour of self-sacrifice',[4] Marx once again owed a great deal to Smith Ordinary labour is a sacrifice according to Smith, it is 'the toil and trouble'[5] he *must* expend [6] It is a giving up of his real desired interest [6] 'He must always lay down the same portion of his case, his liberty and his happiness '[6] In 1765 Smith wrote 'Some reflections on the general behaviour and disposition of the manufacturing population of this kingdom showing by arguments drawn from experience that nothing but necessity will enforce labour '[7] Thus one basis of labourers' alienation was implicit in Smith's initial conception of commercial labour as a sacrifice

Unlike Marx, Smith only implicitly questioned why labour, although it was man's 'fundamental property', only had a value for labourers when it was being sacrificed for actual property goods He never became as explicit as Marx Nevertheless, Smith recognized that the labourer was effectively powerless in modern contracts and factory conditions to prevent himself from being a commodity Thus Smith's ordinary labourer who was already self-estranged by his narrow detail factory work and becoming increasingly isolated because of his inability to sympathize, comprehend, or communicate with his fellows was made unable to escape these other forms of alienation essentially because he was powerless in most contracts Therefore, Smith anticipated all three types of alienation—self-estrangement, isolation, and powerlessness—identified by Marx in his early works.

Columbia University, New York

[1] *W N* ii, p 236

[2] Compare *W N* i, pp 37, 74, 302–3, previously cited, with Karl Marx, *Economic and Philosophical Manuscripts of 1844* (Moscow, 1961), see especially the chapters on the 'Wages of Labour', pp 20–36 and 'Estranged Labour', pp 67–84

[3] Marx, *1844 Manuscripts*, pp 20–8 and pp 67–72

[4] Ibid, p 72 [5] *W N* i p 34 [6] *W N* i, p 37

[7] *Lectures*, p 257 And see *Draft*, p 327, and *W N* i, pp 72–3, and *Lectures*, p 163

A COMMENT ON Y. AKYÜZ: INCOME DISTRIBUTION, VALUE OF CAPITAL, AND TWO NOTIONS OF THE WAGE-PROFIT TRADE-OFF

By KLAUS JAEGER

IN a recent interesting paper Y. Akyuz tries to demonstrate '. . . that there are at least two alternative notions of wage-profit trade-off with different implications for income distribution and the relation between the latter and the value of capital'.[1] To show this, he uses the standard Hicksian two-sector-model and distinguishes between two alternatives: the Hicks case with the price of corn as numéraire and the Sraffa case, where the value of net output (per head) is taken as numéraire. We wish to point out that Akyuz's distinction between the two notions of the wage-profit trade-off is somewhat misleading because it is unnecessary, *if the crucial relations are carefully interpreted.*

If we take the price of corn (π) as numéraire, then—following Akyuz—the price and quantity equations are respectively[2]

$$1 = \alpha rp+\beta w \quad \text{consumption good (corn) industry} \tag{1}$$

$$p = arp+bw \quad \text{capital good (tractor) industry} \tag{2}$$

and

$$1 = bgT+\beta c \quad \text{labour equation} \tag{3}$$

$$T = agT+\alpha c \quad \text{capital per head.} \tag{4}$$

In (1) and (2) w is the wage rate and p the price of capital goods, *both measured in terms of corn.*

The value of net output per head (in corn units) is

$$y = Tpg+c. \tag{5}$$

The well-known Hicks $w-r$ trade-off follows from (1) and (2)[3]

$$w = \frac{(1-ar)}{\beta+rm} = w_H \quad \text{and} \quad m = \alpha b-a\beta. \tag{6}$$

[1] Y Akyuz, 'Income distribution, value of capital, and two notions of the wage-profit trade-off', in *Oxford Economic Papers*, July 1972, vol 24, no 2, p 156

[2] We use the same notation as Akyuz, i e

α, a	Capital coefficients in consumption and capital goods industries
β, b	Labour coefficients in consumption and capital good industries
T	Quantity of capital per man
c	Quantity of consumption per man
r	Rate of profit
g	Growth rate of capital.

[3] Subscripts S and H refer respectively to Sraffa and Hicks's notion of trade-off (see Akyuz)

To derive a relation between the ***share of wages*** in the ***value of net output*** and the rate of profit (the so-called Sraffa case), Akyuz takes another numéraire, i.e. equation (5) with $\pi y = 1$. This, however, is completely unnecessary. Divide (1) and (2) by (5) and define the share of (corn) wages in the (corn) value of net output as w^*, i.e.

$$w^* \equiv w/(Tpg+c) = w/y, \tag{7}$$

then (1) and (2) become
$$1/y = \alpha rp/y + \beta w^*, \tag{1 a}$$

$$p/y = arp/y + bw^* \tag{2 a}$$

From (1 a), (2 a), (3), (4), and (5) it follows that

$$w^* = \frac{(1-ar)(\beta+gm)}{\beta+m(g+r-agr)} = w_S. \tag{8}$$

Equation (8) is exactly identical with Akyuz's equation (7), which he calls the Sraffa wage-profit trade-off. It is clearly different from relation (6), because it relates not the *wage rate* (as (6) does) but the *share* of wages *in net output* to the profit rate. Why should this relation (8) be called a Sraffa wage-profit trade-off? With given g, it simply describes a trade-off between the profit rate and the *share* of (corn) *wages in the* (*corn*) *value of net output* of the Hicksian case. Obviously the two trade-offs (6) and (8) have different implications for income distribution. This is not the case because two different numéraires have been taken but rather because these trade-offs show the well-known fact that the wage *rate* and the wage *share* are two different things.

In our opinion, the distribution of net output between workers and capitalists *is* a matter of social conflict. As the above exposition, however, shows, this is—contrary to Akyuz's statement—completely independent of the chosen numéraire. Within the framework of our (and Akyuz's) model, the only problem of distribution is whether the *wage rate* (in terms of money, corn, labour, or any other units) *or* the *share of wages in net output* should be chosen as appropriate measure for describing the position of the classes in relation to one another. If at all, then the latter seems more adequate.

The price equation (1) and (2) being dual to the quantity equation (3) and (4), the dual relation to (6) is the $c-g$ trade-off, solve (3) and (4) for c

$$c_{S(H)} = \frac{(1-ag)}{(\beta+gm)}. \tag{9}$$

The price of corn (π) being numéraire ($\pi = 1$), the equation (9) describes in the Hicksian case a trade-off between g and both the quantity (c) and the value (πc) of consumption per head; whereas in the Sraffa case, (9) only gives the $c-g$ relation. Therefore, Akyuz calculates another trade-off the

so-called Sraffa $c-g$ trade-off with equation (5) as numéraire ($\pi y = 1$) But this relation (his equation (10)) is again nothing else than a trade-off between g and the ***share of consumption goods*** (corn) ***in net output*** (measured in corn units) of the Hicksian case. This can be shown as follows Divide equations (3) and (4) by (5) and define

$$c^* \equiv c/(Tpg+c) = c/y \tag{10}$$

to obtain

$$1/y = bgT/y+\beta c^*, \tag{3a}$$

$$T/y = agT/y+\alpha c^*. \tag{4a}$$

Solve (1), (2), (3a), (4a), and (5) for c^*

$$c^* = \frac{(1-ag)(\beta+rm)}{\beta+m(g+r-gra)} = \pi_S c_S. \tag{11}$$

In deriving (11) (which is identical with Akyuz s equation (10)), we have not chosen another numéraire, the price of corn still equals unity Therefore, there is no need to distinguish between Hicks and Sraffa trade-offs Notwithstanding, we must distinguish between the following relationships (1) The trade-off on the one hand between g (given r) and consumption per head and (2) on the other hand the trade-off between g and the share of consumption in net output, i e between equations (9) and (11) This, however, is obvious because these equations describe different things

University of Konstanz, W Germany

THE DETERMINANTS OF INTERNATIONAL PRODUCTION[1]

By JOHN H. DUNNING

Introduction

THERE are few branches of economic analysis which are not directly relevant to an understanding of the origin and growth of multinational enterprises (MEs) The subject is obviously of interest to those concerned with the resource allocative activities and financial management of firms, and with the theory of industrial organization Since their operations straddle national boundaries, and involve trade in both goods and factors of production, they come within the scope of international economics, and as vehicles for the transference of new skills and technologies, they are no less pertinent to the theory of economic development. The sharing of the costs and benefits of their activities between the countries in which they operate raises complex and fascinating issues for the welfare economist The geographical flexibility of their procurement, production, and marketing strategies adds a new dimension to the theories of industrial relations and collective bargaining, while their operations are not only influenced by, but help to fashion, a whole range of monetary and fiscal policies used by national governments to advance economic and social goals

I make these observations by way of introduction, because, in interpreting the various explanations of the origin and growth of international business, one is very conscious of the particular interests of the researcher This is shown both in the type of questions asked, and the approach and techniques used to answer them The questions 'why do firms invest overseas?', 'where do firms locate their foreign operations?' and 'what determines the amount and composition of international production?' pose similar, but not identical issues. Each is concerned with the behaviour of firms, but while the first draws on the techniques of micro-investment theory, the second is of interest to the location theorist, and the third needs a knowledge of international trade and industrial organization theory Moreover, each of the questions may be tackled from a positive or a normative viewpoint, and with sectoral, national, or cosmopolitan interests in mind

The purpose of this paper is two-fold, first to survey and critically evaluate the attempts so far made to answer the general question 'why

[1] A shortened version of this paper was presented to a conference on *The Growth of Multinational Enterprises* organized by Gilles Bertin at Rennes, France, in Sept 1972

international direct investment and production?' and, second, to suggest some possible lines for further research, illustrating from data recently published about the operations of U S affiliates in the U K.

The issues involved

What, then, is the subject for explanation? Basically, most writers have been concerned to explain the growth and significance of enterprises which operate and control income-creating activities in more than one country, or, more specifically, the growth and significance of the foreign activities of such companies. It is when one starts to translate this general rubric into operational terms that one runs into difficulties Precisely at what point does an enterprise become 'multinational'? What does one mean by 'control'? What exactly are income-creating activities?

The ME has been variously interpreted in the literature (Aharoni, 1971). Definitions range from those which embrace all firms which operate and control income-creating activities in more than one country (Brooke and Remmers, 1970, Dunning, 1971) to those which would include only those enterprises which operate a common management and operational strategy towards their foreign and domestic operations (Perlmutter, 1969, Behrman, 1969) Others introduce more pragmatic constraints, e g the number of countries in which a firm operates (Vernon, 1972) or the proportion of total sales, assets, or employment accounted for by their foreign activities (Bruck and Lees, 1966). There is also the totally different approach which interprets multinationalism in terms of the geographical spread of ownership or control of equity capital (or capital employed) While respecting the views of the particularists, I have long favoured a broad rather than a narrow definition of the ME, partly because all other definitions are bound to be arbitrary, and partly because I do not consider the attributes of the ME stressed by Perlmutter *et al* are necessarily unique to such enterprises, cf, e g, multi-regional national or international trading enterprises The distinction between the geographical origin of capital and of the ownership of production facilities is best overcome by placing the appropriate adjective between the words 'multinational' and 'enterprise' (Dunning 1971)

Second, as regards the question of control, definitions again vary from including affiliates and associated companies of MEs in which there is any financial stake to those in which there is a 100 per cent equity holding Here, too, there is no purist definition, simply because there is no such definition of control of decision taking, either of its amount or its extent But this much can be said Since, first, a 51 per cent ownership of equity capital ensures the *power* of control over decision-taking, and, second, an overwhelming proportion of the capital of the affiliates in which MEs have

a stake *is* financially controlled by them,[1] we would not go far wrong by considering all companies with a foreign direct investment stake as MEs

Third, the interpretation and measurement of income-creating activities. These include all activities in which there is a capital stake of some kind involved This immediately distinguishes multinational *producing* enterprises from multinational *trading* enterprises. Again, in practice, the line between setting up one's own sales outlet and using a local distributor may be difficult to draw, but this need not greatly concern us, as the great majority of MEs are engaged in the production of goods or financial services and most of the current discussion about their origins and effects is to do with these companies, rather than with wholesaling or retailing ventures

The measurement of the economic activities of MEs raises no new conceptual problems and, in most cases, the indicator chosen will be determined by the data available and the purpose of the exercise In general, output measures are preferable to input measures, for the simple reason the latter are usually expressed in terms of one input, e g capital stock, investment flows, employment, etc Output indices on the other hand, pose the problem of whether output should be *gross*, i e values or sales, or *net*, i e sales less purchases from other firms In fact, most published statistics of international production are of sales rather than net output; these tend to exaggerate the direct economic contribution of MEs or their affiliates to the gross national products of the countries in which they operate

This last point raises two others The first arises when one asks the question 'from whose viewpoint are we measuring income-creating activities?' From the viewpoint of MEs, their own sales, or net output or profits may be the appropriate index From the viewpoint of countries in which they operate, the contribution they make to the gross national product may be the chief (economic) consideration, this includes not only their own output but the effect which they have on the net output of other economic agents in the economy. But the value of this contribution depends on the assumptions made about what would have happened in their absence and, in any case, it may be reasonably argued that since it is individual firms that take the decisions about their activities (albeit in response to signals from governments), it is the factors which influence these decisions which are the relevant ones But in looking at the appropriate policies for governments to pursue towards MEs, the external effects of their behaviour may be equally important

The second point is specific to foreign direct investment and arises because

[1] For example, according to the U S Department of Commerce, in 1966, 95 per cent of the earnings and 93 per cent of the net capital flows of U S foreign affiliates were accounted for by affiliates in which there was a 51 per cent or more U S equity stake Similarly, in 1965 91 per cent of all U K direct investment, outside oil, banking, and insurance, was 51 per cent or more U K financed

the factor inputs of MEs are sourced from both (*a*) the countries in whi they operate and (*b*) other countries. The ratio of (*a*) to (*b*) will determi that part of the value added by a particular affiliate from which the re of the enterprise benefits. This means that, just as expressing the activi of an affiliate in terms of a single input may underestimate its contributio so assessing it in terms of gross or net output may overstate its contributi to the enterprise. If, for example, a ME owns one-half of the equity capital its affiliate, then the ratio of the sales to its affiliates to the total sales of t enterprise will be twice that of the ratio of the profits earned by the affilia to that of the enterprise as a whole, assuming that the profit/sales rati and taxation rates are the same for all the operating units of the enterpris Once again, it depends on what questions one is seeking to answer, but is worth emphasising that identifying and measuring activities of MEs not as straightforward as it may appear to be

In practice, the matter is often settled by the data available and tl economist has to cut his coat according to the cloth given him, or obtain by himself! And the research so far done on the growth of the multination enterprise strongly reflects this constraint. Broadly speaking, economis have obtained their data from three sources First, from information pu lished, mainly by governments of host and investing countries, on the sto or flow of inward and outward direct investment Due mainly to differe reporting requirements of governments, the form, coverage, and reliabili of these data vary enormously between countries, and within a countr over time, and they are rarely directly comparable. Valiant attempts ha been made by Polk (1971), Behrman (1969), and Rolfe (1969), to constru a world matrix of the value of international direct investment and/ production, but none has been completely successful [1] As far as *investir* countries are concerned, the most comprehensive data are those publishe by the U S Department of Commerce [2] These include investment, outpu and income data for 1966 and 1970, broken down both by country an industry U K statistics are confined to a fairly detailed geographic breakdown of the capital stake and investment flows, the industri breakdown is very broad There are also quite reasonable data on outwar direct investment for Canada, Australia, and latterly Japan. Of the *ho* countries, Canada, Australia, the U S , and U K , Sweden, and Belgiun and some of the l d c s , e g Argentina, India, Ghana, Nigeria, Malaysi

[1] The work of Judd Polk deserves mention in this context He defines (U S.) productio abroad as 'production in which U S management and financing work together with foreig factors of production' (Polk, 1971, p 9), or, even more succinctly, 'product emanating fro foreign investment' or the 'product profits of an investment activity abroad' His estimat of the value of this component of world production in 1969 was $450 million, or 15 per ce of gross world product, and that since 1950 this has been increasing at a steady rate of ! per cent a year (Polk, 1971, pp 5 and 8)

[2] Usually in the *Survey of Current Business* or in special supplements to this periodica

Korea, the Phillipines, Taiwan, and Indonesia provide reasonably good statistics.[1] Major surveys on the extent and pattern of Swedish and German foreign investments are currently being undertaken

The second form of data is that derived from field work carried out by research institutions or individual research workers in pursuance of a specific project to do with foreign direct investment or the ME Most of these projects have taken the form of country or industry case studies,[2] although some have been specifically concerned with the determinants of foreign investment. Again, the quality of the data varies as, in most cases the investigators have had to rely on the good offices of firms, but for some of the less-well-documented countries, particularly the l d c s., and for a more detailed breakdown of industry statistics, these studies usefully supplement (and sometimes improve upon) the official statistics

The third source of information is that being gradually amassed in data banks and is based largely on statistics related to individual companies. The first of these was established at Harvard and supplied much of the data for the studies led by Raymond Vernon (1972), more recently, Gilles Bertin has set up a European counterpart at Rennes In spite of the interpretative difficulties, I believe that these data banks have much to commend them. Already, useful progress has been made by international and government agencies, notably U N C T A D , O E C D , and E E C , and U S Tariff Commission, by research institutions, e g Foreign Policy Research Institute at Pennsylvania and Centre for Multinational Studies at Washington and by such organizations as the International Chamber of Commerce and Business International Various international trade union secretariats are also actively gathering material One feels that the time is rapidly approaching for some rationalization of data collection, partly to avoid unnecessary duplication of research and clerical effort, and partly to reduce the work on individual MEs, who, after all, are the main providers of information.

I will return to this point later in the paper But one practical difficulty should be mentioned here The number of enterprises which make up the great bulk of foreign direct investment is small The largest fifty MEs probably account for one-half of the total international direct investment and an even higher percentage of international production in the world the next largest fifty account for up to another 25 per cent When one comes to break down these operations geographically and industrially in any meaningful sense one is soon dealing with a handful of companies The possibility of identification then becomes very real, and this may well

[1] For a comprehensive analysis of foreign direct investment in Asia and the Far East see United Nations, 1971

[2] See particularly those mentioned on p 21, May and Arena (1971), FIEL (1971), and United Nations (1971)

impose the ultimate limit to sophisticated econometric work in this fiel
This point is further underlined when one comes to classify MEs by tl
operational strategies they pursue, and/or other variables, e.g. by activit
intra-group exports, size, age, etc ; one is soon down to a few observatio
in each cell of the matrix which raises conceptual as well as identificatic
problems.

So much by way of introduction We now turn to examine the work
far done on identifying the factors influencing the origin and growth
international production We shall mainly concentrate on the *positi*
approaches to the subject and discuss these under six main headings.

1. The survey approach

One approach to explaining the extent and character of foreign busine
operations has been to ask the companies themselves to identify the reaso
for their behaviour. Usually, this approach has confined itself to analysir
the *initial* decision to produce abroad, and, more often than not, tl
questions have been formulated in the most general terms, e g 'what a
the main factors which influenced your decision to invest overseas?', an
rarely does any guidance seem to have been given to the respondents as t
assumptions underlying the questions asked Because of this, the survey
have produced a wide range of answers, which reflect as much the respo
dents' interpretation of the questions as the determinants of the inves
ment decision

There were several surveys of this kind in the later 1950s and 196
(Barlow and Wender 1955; National Industrial Conference Board, 196
Robinson, 1961; McGraw Hill, 1961; Behrman, 1962; Basi, 1966, Hakan
1966, Kreinin, 1967; Kolde, 1968; Hogan, 1968), and frequently, too, i
broader based works on foreign direct investment (Safarian, 1966; Brasl
1966, Brooke and Remmers, 1970, Deane, 1970, Daniels, 1971, Andrews
1972, Forsyth, 1972), questions of this type have been asked. Some o
these focused on the goals of foreign direct investment, and others on mean
of achieving goals, but most did not distinguish between the two In th
main, the results of the surveys were presented as a tabulation of the reason
for moving abroad, or to particular countries listed by the respondents in th
sample, in Basi's analysis, a three-point 'importance' scale was used, bu
mostly the only evaluation was by the times particular determinants wer
mentioned, the number of which ranged from nine in the Kolde study t
twenty-five in the Robinson study No attempt was made to classify th
results by types of economic activity, or by country; although some of th
studies concentrated on particular regions within countries or areas (John
and Brash (Australia), Forsyth (Scotland), Kreinin (Europe), Hakan
(Nigeria))

It is clear that these studies can, at best, do little more than identify and perhaps rank by importance the sort of factors which businesses take into account in establishing production units abroad. At worst, they can be thoroughly misleading. Quite apart from the confusion between goals (e.g increased profits or share of market) and factors affecting the achievements of goals (e g. transport costs, market growth, etc) the reasons cited by firms were sometimes interdependent on each other, e g lower costs of production and higher labour productivity, in some cases the reasons cited were quite specific, e g. the existence of local engineering facilities, or to match a rival's investment, in others, they were very general, e g diversification, inflation. Moreover, as we have said, there was little attempt to classify types of foreign operations, and only a casual acquaintance with the literature suggests that the determinants of investment vary so much with the *type* of investment, cf the reasons for upstream and downstream investment, that any generalizations are not very helpful.

As Table I illustrates, almost without exception, the studies stress the host government's attitude to inward foreign investment, political stability, and the prospects of market growth as the most important considerations prompting foreign activities, next in order come the fear of losing an existing market, the likelihood of exchange rate fluctuations, limitations imposed on foreign ownership, and barriers to trade [1] Only a minority of firms appear to have been enticed abroad by lower production costs, neither do savings in movement costs loom large in their calculations But again, the studies do not tell us the way in which these determinants may vary with geographical or industrial composition of the investment In summary, they may be criticized, partly because they fail to differentiate between motives and determinants, partly because they do not identify the assumptions underlying the answers given by firms, and partly because no attempt is made to normalize for differences in the characteristics of firms (or countries) Certainly none of them take us much further in a generalized theory of international production, or help us to understand the determinants of new investment once the initial locational decision has been made

More recently, efforts have been made to improve the methodology of the survey approach, both by giving respondents a clearer conception of the type of variables it is sought to identify, and by suggesting ways in which they might be evaluated.

Stobaugh (1969*a*) for example, makes use of a matrix which identifies two main groups of variables which gauge locational attractions to companies For example, he relates *product-related* influences, technological and

[1] For an interesting examination of the reasons for establishing foreign manufacturing plants by U S. firms prior to 1900, see Vernon (1972) (Table 3 5, pp. 72–3) and Wilkins (1970).

TABLE I ***Summary of determinants of foreign direct investment (selected studies)***

Number of times factors mentioned

	(a) *Foreign investment in general*					(b) *Investment in specific countries*			
Name of researcher Date of publication Number of firms in sample	Robinson[1] (1961) 205	Behrman (1962) 72	Basi[2] (1966) 214	Kolde (1968) 104	Forsyth[(a)3] (1972) 105	Brash (1966) 100	Deane (1970) 139	Forsyth[(b)3] (1972) 105	Andrews[4] (1972) 80
(a) *Marketing factors*									
(i) Size of market			141	.					
(ii) Market growth	262	19	158	7	82	89	21	14	28
Note: Robinson 262 and Deane 21 are bracketed across (i) and (ii)									
(iii) To maintain share of market or match a rival's investment	130	.	126	12	35		30	6	
(iv) To advance exports of parent company		1			2	.	..	1	.
(v) Necessity to maintain close contact with customers	..	7	..	..	5	..	15	9	
(vi) Dissatisfaction with existing market arrangements	..	3	..	25	..	..	..	..	..
(vii) Export base for neighbouring markets	104	3		..	..	30		.	39
	496	33	425	44	124	119	66	30	57
(b) *Barriers to trade*									
(i) Barriers to trade	130		..	21	28	78	76	..	11
(ii) Preference of local customers for local products		14			1	24	.	.	
Note: Behrman 14 is bracketed across (i) and (ii)									
	130	14	..	21	29	102	76	..	11
(c) *Cost factors*									
(i) To be near source of supply	.	..	.	..	3	.	14	2	..
(ii) Availability of labour	209		*					53	.
(iii) Availability of raw materials		12	114		..		7		.
(iv) Availability of capital/technology	.		78			.		11	
(v) Lower labour costs	79	..	103	.	.	.	.	18	40
(vi) Lower other production costs		7		20		11	.		
(vii) Lower transport costs					.	22		18	
Note: Forsyth(b) 11, 18 and 18 are bracketed together, with Andrews 40 beside the bracket									
(viii) Financial (*et al*) inducements by governments	50	..	..	..	1	13	..	52	45
(ix) General cost levels more favourable (less inflation)			134		.		14		

(d) Investment climate									
(i) General attitude to foreign investment	[5]	.	145	6	..	..	10	..	.
(ii) Political stability	115	..	159		..	.	.		..
(iii) Limitation on ownership	20			.	.		.	..	.
(iv) Currency exchange regulations	105[6]				.	.	.		.
(v) Stability of foreign exchange			151						
(vi) Tax structure			131	4	.		.		
(vii) Familiarity with country			100						
	240	..	686	10	..	—	10	..	..
(e) General									
(ii) Expected higher profits	182	20	144						
(ii) Other[7]	252	14	112	5	14	37	39	43	50[8]
	434	34	256	5	14	37	39	43	50
	1638	97	1796	100	171	304	226	227	203

* Included in lower labour costs
1 Number of times factors are ranked 1–3 in a 6-point scale
2 Listed as 'crucially' or 'fairly important' in Basi's 3-point scale
3 Forsyth[(8)] refers to reasons given by firms on decision to invest outside the U S
4 Andrews' survey was concerned with identifying reasons for investing in Ireland
5 Dealt with in a separate part of the survey and regarded as crucially important
6 Classified as 'financial stability'
7 Including 192 mentions for availability of infrastructure, power, and banking facilities
8 Including forty mentions 'to take advantage of Ireland's entry into the Common Market should that occur'

marketing characteristics, life cycle pattern, cost structure, and economies of scale to *country-related* influences, e.g. market size, investment climate, local technology, and distance from major exporting nations Schöllhammer (1972) adds a third group of influences, viz. *company-related* influences, e.g. size of firm, scope of international operations management strategy.

The same authors and Piper (1971) have also suggested schemes for the evaluation of these variables. Stobaugh (1969*b*), for example, sets out ranges of marks which might be given for each particular environmental variable (attitude to capital repatriation (0–12), extent to which foreign ownership is allowed (0–12), currency stability (4–20), etc), which are then assigned by firms according to some predefined criteria The marks are then aggregated and an index of environmental attraction, or investment climate, obtained Schöllhammer (1972) in a study of 140 American and European MEs asked corporate executives involved in making location decisions to rank seventy-eight country-related influences (classified into nine broad categories, e g. economic, legal, geographical, political, labour, tax, etc , factors) on a scale from 1 (of no importance) to 4 (very important) His findings broadly confirmed those of earlier surveys. The two most important individual location factors were existing market size and anticipated market growth, but of the nine broad groupings, political, supply, and tax considerations outranked the rest

Such schemes as these have their obvious attractions, but they also have their drawbacks; among the latter are first they almost always set the same standards for all types of investment and in all countries; second they assume that over the life of the plant, the investment climate will remain unchanged, and third they assume that individual locational determinants can be separately and independently evaluated

We conclude while the survey approach may be helpful in identifying the factors which influence international production, it can do little more than this In the past, it has not been satisfactory in evaluating particular goals or determinants, and even attempts to use a ranking procedure have been of limited value because of the failure to take account of different types of investment None of the surveys have so far distinguished between factors affecting the establishment of foreign production units from those influencing increases in international production Finally, all too frequently they have rarely defined the form of the involvement of companies abroad (when investment is taken as the dependent variable it is not clear whether this means investment-owned or investment-controlled).

2. Capital theory

The second approach to the study of 'why international production?' focuses attention on one factor input, viz. capital, or changes in capital,

viz. investment, and is essentially an extension of received capital theory Mainly because of data constraints, almost all the empirical work done in this area has been on the behaviour of U S MEs In most cases, the U S. *share* of the capital stock, or investment, of U.S. foreign affiliates is taken as the dependent variable, but occasionally the *total* plant and equipment expenditure of affiliates, i e investment in fixed assets, is used

The traditional theory of international capital movements asserts that such movements arise because of differences in the levels of interest rates between countries Under these conditions, money capital flows across the exchanges, if the margin by which the expected yield exceeds the cost of capital is greater than that of projects at home Until the mid 1960s, this relationship was thought, by most economists, to explain movements in *portfolio* investment fairly well (Mundell, 1960, Kenen, 1963) Since then, partly as a result of developments in the theories of investment behaviour and portfolio distribution, a new view has emerged which argues that, while the allocation of the *stock* of assets held at home and abroad depends on the level of interest rates and risk evaluations, changes in this allocation, i e capital flows, will depend on *changes* in interest rates (Branson, 1970, Floyd, 1969) According to this view, an increase in foreign interest rates will have a two-fold effect First, it will cause a shift in the stock of portfolios towards foreign assets, this is called the so-called *stock-shift* effect, which will vary *inter alia* with the size of the portfolio and the amount of change in the interest differential Second, there will be a reallocation of portfolios at the margin towards foreign assets—the so-called 'continuing flow' effect Where the latter component is small, the supply-elasticity of capital with respect to changes in interest rates between countries is likely to be substantial only in the adjustment period For there to be a permanent redistribution of capital movements between countries their relative interest rates must be constantly changing This new view is generally supported by the empirical studies of the last few years (Branson and Hill, 1971), although the period of stock adjustment is now being shown to be somewhat longer than was first thought

It is generally accepted that models of this kind, designed to explain international flows in portfolio investment, can only partially explain the international capital formation of firms or that part of it financed by direct foreign investment This is mainly because, unlike movements in portfolio capital, which are essentially financial transactions between independent lenders and borrowers, direct investment involves no change in ownership It does, however, involve the transmission of other factor inputs than money capital, viz entrepreneurship, technology, and management expertise, and is likely to be as affected by the relative profitability of the use of these resources in different countries as that of money capital

(Stubenitsky, 1970). Put another way, the models are inadequate because they assume that the transactors engaged in the activity of international investment have similar behavioural characteristics (Learner and Stern, 1972).

Nevertheless, recent research on the origin of international financial and real capital flows has provided useful new insights which have a direct bearing on the investment behaviour of MEs (Spitaller, 1971; Stevens, 1972) Harry Johnson (1966), for example, makes the useful distinction between movements in capital which occur in response to interest rate differentials and those which are generated by the expectancy of higher profits At a macro-level, it is this latter type of movement that Borts and Kopecky (1972) argue can best be explained by the same factors which explain economic growth, e g increases in population, technological advances, the improvement in the terms of trade between exports and imported capital goods, the savings rate, and the capital coefficient; and that it is not normally necessary to introduce monetary factors to explain why or how capital transfers occur. Monetary variables may affect capital movements but only in so far as an excess demand for liquid assets has an influence on the excess demand for goods

The alternative approach is more *micro*-oriented and represents the main stream of thinking on the subject This is directed to extending the theory of domestic corporate investment to the international activities of firms. There are two main strands to this approach The least well developed is that which looks at the firm's foreign investment decision as an extension of the theory of portfolio distribution. Following attempts by Grubel (1968), Miller and Whitman (1970), and Levy and Sarnat (1970), to explain the distribution of *portfolio* investment across national boundaries using a stock adjustment model of the Markowitz (1959)/Tobin (1958, 1965) variety, Prachowny (1972) and Stevens (1969*a*) set out to test whether or not firms allocated their *direct* investment expenditures so as to maximize a utility function positively related to expected returns and negatively related to risk Their results were inconclusive, particularly when disaggregated data were used Cohen (1972), on the other hand, has demonstrated that large U S. corporations with more extensive foreign activities tended to have smaller fluctuations in their profits during the 1960s Finally Mellors (1973), using a technique first developed by Smith and Schreiner (1969) to explain the domestic diversification of conglomerate firms, has demonstrated that the geographical allocation of direct investment by U K firms, in response to *post*-tax rates of return, provides some support to the portfolio model.

More extensive have been the attempts to apply various models of domestic capital formation by businesses to explain foreign investment [1]

[1] For a survey of some of the recent literature, see Stevens (1973).

In particular, two main lines of research may be mentioned. The first is an extension of the neo-classical theory of real investment, and assumes the maximization of the market value of assets to be the goal of firms. Here the most popular model is that developed by Jorgenson (1963), in which investment is viewed as a gradual adjustment of a firm's actual capital stock to its desired level, i.e. $K_t^* = ap_t Q_t/c_t$ where K_t^* = desired level of capital stock (at time t), p_t = product price (at time t), Q_t is expected output, c_t = the rental price of capital (which in turn is a function of the price of capital goods and its rate of change, the cost of capital, the depreciation and tax rates), and 'a' is a constant from the Cobb–Douglas production function measuring the elasticity of output with respect to capital. This is a modified version of the flexible accelerator explanation of investment, which in most tests has out-performed the simple accelerator model, liquidity and cash flow models, and security valuation models (Stevens, 1972).

There have been numerous studies which have examined the determinants of foreign investment over the last five years. Again, it is convenient to classify these into two groups. The first is illustrated by the work of Stevens (1969*a* and 1972), Moose (1968), Severn (1972), Popkin (1965), Kopits (1972), Richardson (1971 and 1972), and Kwack (1973). Each of these strongly supports the standard investment theory by demonstrating that expenditure by U.S. firms on foreign plant and equipment is highly correlated either with the sales of U.S. foreign affiliates or some measure of output for the area of industry in question. Severn, for example, used a two-country model (the U.S. and the rest of the world) to explain differences both in the specification of domestic and foreign investment functions, and the distribution of corporate funds between home and foreign uses. He concluded that subject to a liquidity constraint, investment was strongly correlated to changes in sales in both cases. He also asserted that MEs allocated funds without reference to national boundaries and that, eliminating factors common to both foreign and domestic investment, the two were at least partially substitutable and interrelated through the financing mechanism. Popkin, in his study of U.S. manufacturing affiliates (1965), claimed that the relative profit rates and other financial variables were more important than market structure or technological factors in explaining variations in the behaviour of firms. Stevens (1972), using similar data, and an extension of the Modigliani–Miller theorem (1958), derived equations which, *inter alia*, related plant and equipment expenditure and changes in current assets to the present market value of firms, and also financial flows to the same goal and that of exchange loss minimization. He found that all equations explained past data quite well. In his study, Kwack (1973) identified a negative correlation between U.S. interest rates and foreign

investment but, like Stevens (1972), concluded that the voluntary restraint programme, aimed at improving the U.S balance of payments was statistically insignificant

Richardson (1971 and 1972) took these discussions a stage further by distinguishing between different types of foreign direct investment. In particular, he argued the need for a separate theory to explain investment in new ventures, the main goal of which is likely to be market penetration, rather than the profit-maximizing or growth goals of established ventures. He also considered that domestic-type theories were less successful in explaining the investment policies of the affiliates of integrated multinational firms, which were more likely to be geared to a global strategy, than in the case of independent affiliates, where an 'every tub on its own bottom' type of policy was the usual practice In his contributions, he suggested the kind of modifications necessary to the accepted variables to explain the optimal capital stock of each of these types of foreign investment, although he did not attempt to put these to the test His, however, is perhaps one of the most rewarding lines of research in that he recognizes both the motives and determinants of MEs will vary according to the type of foreign operation, a point to which we shall return later

Most of the research so far mentioned accepts that there are certain factors affecting foreign capital formation which are specific to such investment, although Herring and Willett (1973) have demonstrated that between 1957 and 1969, U S plant and equipment expenditures at home and abroad were significantly correlated Other studies have attempted to isolate some of these, e g Cairncross (1973) and Herring and Willett (1972) control over capital exports, Kopits (1972) and Mellors (1973) the tax variable, Stevens (1969*b*) and Heckerman (1969) exchange risks, and Horst (1972) and Jud (1973) tariffs Much more difficult is to test statistically the significance of non-quantifiable variables such as the investment climate, which, as we have seen (pp 296–7) businessmen consider to be an influence on their investment plans One way out of this dilemma has been suggested by Miller and Weigel (1972) who argue, on the lines of Aharoni (1966), that decisions about the location of investment should be regarded as a two-stage discriminant process In the first stage, locations are classified as 'suitable' or 'potentially unsuitable', on fairly basic grounds, size of market, prior investment, barriers to exports, investment climate, etc In the second stage detailed calculations are made of the expected economic profitability of the locations considered potentially suitable The fact that the variables included in many models may not hold up well when predicting *all* decisions may be due to the fact that the first stage rejections have already been made, but on non-economic grounds.

The second group of studies on investment behaviour have sought to

explain movements in foreign capital formation in particular geographical areas; the writings of Bandera and White (1968), d'Arge (1969), Scaperlanda (1967), Wallis (1968), Scaperlanda and Mauer (1969, 1971, 1972), Schmitz and Bieri (1972), all dealing with U.S. direct investment in Western Europe,[1] are some examples. Most of these, using either time series or cross-sectional data, relate absolute amounts of investment (or capital stake), or shares of investment (or capital stake) to profit rates, size of markets, growth of markets, tariff rates, and some kind of trend and/or slope shifting variables; the 1968 Bandera and White study included an international liquidity variable. The cross-sectional studies strongly support the hypothesis that U S investment has been directed most to countries with the fastest rate of growth of GNP, with profitability and other variables, including tariffs, being a secondary consideration On the other hand, there is little evidence that, in itself, the formation of the E E C. had a substantial effect on the level or direction of U S investment flows; although much depends on the precise specification of the relationships, the level of disaggregation,[2] and the years for which the comparison is made The time series data lend support to the cross-sectional data when the capital stake is taken as the dependent variable (Bandera and White, 1968) Again, in both cases, the market variable showed up better than the profit rate

What conclusions can be drawn from these studies? In my view, none of them can take us far along the way to understanding 'why international production?' This, I think, is chiefly because their attempts to explain either foreign capital formation or movements in capital across national boundaries evade the more interesting questions to do with international production. The studies take, as given, the value of variables which, themselves, need explaining. Anticipated profits are a good illustration These are almost always expressed in terms of the profitability *of the foreign affiliates* But not only may these be a very imperfect indication of the contribution of the affiliate to the investing enterprise (Reddaway, 1968 and Vaitsos, 1972) but to explain direct foreign investment in terms of profitability begs the question '*why* that profitability?', the answer to which is bound up *inter alia* with the competitive position of foreign affiliates *vis-à-vis* indigenous firms and exports

In other words, the questions asked by capital theory do not get to the heart of the matter The concern of this approach is not to explain foreign investment or capital formation *per se*, but that, assuming this to exist, how far is its allocation influenced by profit rates and market growth? This is

[1] For an analysis of the determinants of foreign direct investment in the U.S see Daniels (1972).

[2] Compare the significance of the tariff variable with that suggested by individual industry studies

a perfectly valid and legitimate interest, and when related to the same variable for domestic investment, does point to some interesting differences in the behaviour of firms (Richardson, 1971). But we are little wiser in understanding *why* this is so

Moreover, all the studies are, to some extent, deficient in their choice of explanatory variables and we have said that investment, particularly investment financed by a particular source, is not necessarily a good index of the activity of firms, as it underestimates the importance of labour-intensive firms Of the independent variables, the rate of profit earned by affiliates may inadequately express their contribution to the organization of which they are part, particularly where there is a good deal of product or process specialization between affiliates, and intra-group trading at other than arm's-length prices Moreover, the more vertically or horizontally integrated an ME becomes, the less significance can be attached to the market size or potential of the country in which production is located This especially applies to those enterprises which practise a policy of global or regional, horizontal or vertical specialization, e g the Philips and I B M s of this world They are not primarily dependent on markets of the countries in which they operate because their decisions will be influenced by other considerations

Finally, the data on which the analyses of investment are based are rarely disaggregated by type of economic activity Because of this, it is impossible to assess the extent to which different types of overseas operations are influenced by different variables This, in fact, one knows to be the case and the importance of the profit (or sales) of the affiliate as the contribution to the goals of the enterprise may be small The objective of foreign sales and marketing ventures is primarily to advance the exports of the investing company; even a manufacturing affiliate may spur the exports of related goods from its parent company (In 1966, U S affiliates imported goods worth $6·1 billion from the U S —about 11 per cent of their total foreign sales), similarly an investment in raw materials may be made to safeguard supplies to the rest of the organization; while Reddaway (1968) and Dunning (1970) have observed that the feedback of knowledge resulting from an investment in a technologically advanced country can more than compensate for any low profits earned.

3. The trade approach

The third approach to 'why international production?' is that of international economics, and stems fom the dissatisfaction with received theory to explain recent trends in the level and composition of trade. It is worth

emphasizing that, in the classical model of static comparative advantage, there is no room for the ME at all. With completely free movement of goods but immobility of factors of production, and with all firms transacting goods and services in a price-taking situation, there is little incentive for international direct investment (Kindleberger 1969). But, with production by firms outside their national boundaries now thought to account for 15 per cent of the world's output, these are no longer reasonable assumptions Standard theory, whether it be of the classical or neo-classical variety, makes no allowance for trade in factor inputs (Baldwin, 1970), largely because the conditions necessary to such trade are assumed not to exist.

The most powerful attempts to incorporate capital movements into trade theory in recent years have come from two directions. First, Mundell (1957), using the Hekkscher–Ohlin–Samuelson model, asserted the proposition that trade and capital movements are substitutes for each other and that the equalization of factor–price ratios implies the equalization of commodity price ratios Second there has been the attempt to take account of changes in technology or advances in knowledge, in the analysis (Johnson, 1968) In the static model, innovations are ignored altogether as production functions are assumed constant and identical (or nearly identical) throughout the world Where they are introduced, e g in a comparative static situation, their benefits are assumed to be instantaneously and freely transferable Such an assumption is totally unrealistic in a situation where information is costly to produce, is enterprise specific, and is sold under conditions of imperfect competition; where governments both finance the output of new knowledge and impose barriers on its dissemination, e g by the patent system, and, hence, affect the patterns of trade and resource allocation.

These and other market constraints both help fashion the initial location of new products and processes, and at the same time, induce the means by which barriers to the diffusion of the knowledge giving rise to these products and processes may be overcome Beginning with Posner (1961), a steady stream of writers have attempted to demonstrate how innovations in one country may affect the comparative advantage of countries, and how the trade initially generated might be gradually eliminated by the recognition and imitation of the innovations elsewhere Various models have sought to explain the process of the transference of production from the innovating country Of these, the product cycle model (Hufbauer 1966; Vernon, 1966; Stobaugh, 1968; Wells, 1972) has, perhaps, come nearest to explicitly recognizing the role of MEs in this process, although in their writings, Hirsch (1967), Wilkinson (1968), Quinn (1970) also accept that it may play an important role Other economists, e.g. Gruber, Mehta, and Vernon

(1967) and Keesing (1966), have also observed the relationship between the production of knowledge, international investment and trade, and more recently Baldwin (1970) has called for an explicit incorporation of trade in factor inputs into trade theory

In explaining how and why international production arises, trade economists have tended to emphasize, first, the conditions under which the foreign markets of a particular country are best exploited through the affiliates of its firms producing in those markets rather than by exports, and second, the possible consequences of this on existing production outlets and trade patterns. The product cycle theory asserts that initially production will be located in the country of innovation, and sold there Exports follow as new markets are sought, but in due course, depending on relative exchange rates[1] and demand and supply conditions in importing countries, indigenous production may become profitable. Whether or not this output will be supplied by local firms or affiliates of firms of the innovating country will depend on the barriers to entry facing the two groups of firms, and their relative efficiencies. It will also be influenced by the strategy of enterprises towards their foreign operations and the type of market structure(s) in which they are competing. Of the barriers to entry facing indigenous firms, Hufbauer (1966) has stressed the technological gap caused by the lag in the international transfer in technology while Vernon (1966) places rather more emphasis on market constraints Both writers, however, see the ME as an instrument for surmounting these barriers. The final phase of the cycle is where these producers may themselves begin exporting, competing with the product-innovating firms even in their domestic markets

Both approaches and other neo-technological theories of trade (Hufbauer, 1971) are micro-oriented and differ from received theory in two major respects First, they are more concerned with the behaviour of firms than of countries; second, as they have so far been presented, they are particular, rather than general models as they tend to endow innovating firms and countries with special economic characteristics and, in consequence, patterns of production and trade. Vernon himself accepts that the product cycle sequence is less satisfactory in explaining the territorial distribution of production of MEs which adopt a global strategy towards their operations (Vernon, 1972), while there is some doubt that the process adequately explains the sequence of events when innovations originate from countries with (relatively) low incomes and wage costs and/or small markets (Dunning, 1971). Nevertheless, the models are of especial interest

[1] Where, for example, the exchange rate of the exporting country is overvalued, outward investment will be favoured relative to exports, where the exchange rate of the importing country is overvalued, inward investment will be favoured relative to imports Some commentators have argued that the rapid increase in the level of U.S foreign investment in the 1960s was due largely to reflected the overvaluation of the dollar.

in that they emphasize the role of innovations in forging new trade patterns within an imperfectly competitive environment, conditions which are the seed-bed of growth of the modern ME.

The value of the trade approach to understanding 'why international production?' is that it reminds us that foreign direct investment is one of various ways of exploiting a foreign market. Though the precise relationship between these alternatives has yet to be analysed in the literature, various case studies of the product cycle (Wells, 1972*a*) give useful glimpses at which point one means will tend to replace another It is, of course, here the overlap with the location theory approach is best seen, it is here, too, where trade theory most needs restructuring to incorporate movements in factor inputs, even if some of these movements are not strictly trade, but the transference of resources between one part of an ME and another

One foresees further interest of trade theorists in this area because of the growing impact of the ME on the flows of goods and services across national boundaries (Robertson, 1971) This will force attention on explaining the behaviour of such companies. Until recently, there have been few data to test systematically any trade-type hypotheses These are now starting to emerge—for example data recently collected by the U S Tariff Commission give sales and exports (including intra-group exports) of U.S. foreign affiliates, classified both by industry and country; although, compared with the quality and detail of trade statistics, the situation is still unsatisfactory. Conceptually, the most helpful lines of approach seem likely to be two-fold; first, the implications of dynamic comparative advantage (Bruno, 1970, Klein, 1973) and particularly those which arise from the incorporation of human capital and productive knowledge into capital (Johnson, 1968); second, an analysis of the way in which trade is influenced by the ways in which markets are exploited (e g by investment, exports, licensing, etc) The explanation of factor endowment and its impact on trade may give some insight into the geographical and industrial composition of international investment Although these are largely macro-concepts, they should provide a useful insight into reasons behind these forms of international transactions.

So far, we have thought of the ME as an instrument for exploiting foreign markets The second aspect of trade theory concerns the question of the distinctive character of the ME as an owner of resources in different countries compared with national (i e indigenous) firms The implications of this will depend on the type of operations and the strategies adopted by the ME, but in principle they raise two issues First, there is likely to be more specialization and integration when such advantages occur than if these operations were conducted by independent firms; second, all

intra-group transactions involve the setting of transfer prices, which v
affect the *terms* of trade. The determination of such prices will depend
circumstances which, again, may require modifications to the assumptio
underlying trade theory.

4. Location theory

Like trade theory, location theory has so far had little to contribute t
an explanation of the level or composition of international production Th
is because location theory has traditionally confined its attention to tl
territorial allocation of resources, and trade of firms *within* national bou
daries, and only Ohlin (1967) and, to a lesser extent, Giersch (1950) an
Weber (1958) have attempted to go beyond this point Yet, removing th
geographical constraint, the theory of location would seem central t
answering the question 'why international production?' Assuming tl
goals of enterprises are unaffected by the countries in which they produc
there is no reason why a U S firm, in choosing between a New York c
a Paris location for its new plant, will be influenced by different criteri
To be sure, additional factors will affect the choice of a foreign locatio
e g the possibility of exchange fluctuations and differences in corporat
tax rates, the risk of expropriation, etc , but, conceptually, there is n
difficulty in embracing these in the basic analysis [1]

Location theory is concerned with both *supply* and *demand* oriente
variables influencing the spatial distribution of production processe
research and development, and administration of firms, unlike trade theor
it is not concerned with the division of labour between countries Assumin
a certain size and distribution of markets, and that each firm is a profi
maximizer operating in a price-taking situation, production will be locate
where costs are lowest (Greenhut, 1952) This, in turn, will depend on th
availability and cost of factor inputs, the efficiency at which these ar
transformed into outputs, and the costs of movement from the point o
production to that of marketing Some of the special features of producin
outside national boundaries can be incorporated into this kind of mode
and an optimal solution found But others may be more difficult to dea
with, e g the possibility of exchange rate adjustments or political actions
partly because they cannot easily be quantified and partly because of th
inherent uncertainty attached to them

[1] Location theory also links with the theory of the growth of the firm Firms expan
either by selling more of the same product to existing markets, or by diversifying thei
products, processes, or markets The territorial spread of production across nationa
boundaries partly arises from a similar diversification of markets, but it may also be linke
to the new opportunities for spatial specialization arising from the diversified product o
process (or even functional) structure of firms

In contrast to this approach, demand oriented theories assume production costs to be independent of location and assert that the distribution of markets and the location of competitors will govern the siting of production units (Losch, 1954) The theories of spatial interdependence are essentially an extension of the principles of monopolistic competition and oligopoly. Each location guarantees an element of spatial monopoly, the extent of which will depend on the character of the market, the locational strategy and efficiency of competitors and movement costs It will also be affected by the character of production, for example whether or not a firm is operating under economies of scale, as this will influence both the extent to which firms tend to cluster or disperse and the number of firms involved

It is now generally accepted that any comprehensive theory of location must incorporate both cost and market factors, and that in an imperfectly competitive situation, the maximum profit location will not necessarily be the one where costs are minimized (Greenhut, 1952) (An analogy is with output and price determination of a firm producing under conditions of oligopoly where it can affect its profits by the size of the market it chooses to exploit as well as its cost conditions)

Again, evaluating these factors as they affect the location of international production, the picture is more complex but not significantly changed. Looked at from the viewpoint of supplying any given market, the question can be discussed in two parts One is to explain the spatial distribution of production units irrespective of ownership; the other is to explain the ownership of these units. Assume, for example, a certain size of market for a particular product in the U K , and that there are two nationalities of firms—U.K. and U S —which could supply that market, then what determines (*a*) the extent to which the market is supplied from production units located in the U K or the U S , irrespective of ownership, and (*b*) given the location of production, whether the nationality of ownership of these units is U K or U S ?

The answer to these questions, to my mind, provides one of the keys to the unique character of the ME and lies at the core of the industrial structure approach to 'why international production ?' For, rephrased, the question asks 'why is a market of a particular country served by the affiliates of foreign-owned firms producing in that country rather than by indigenous firms?

Location theory tackles this question from the viewpoint of individual firms, like capital and trade theory, however, it takes as data the information on costs and market size and structure And, as we have suggested, given this data, it can not only explain actual location patterns, but can also indicate optimal patterns, subject to the uncertainties surrounding particular markets and future events. From the *supply* side, an ME is

faced with the same type of cost decisions as a national enterprise; but i purchasing and marketing options may be wider, and the evaluation · foreign investment climates may be a complicated business (Stobaug 1969). From the *demand* side, one observes the structure of competitio and hence markets served, may be somewhat different The Vernon thes argues that the production of many new products and processes, fir discovered in one country, is later transferred to another by a variety means, one of which is through affiliates of the innovating firms Th assumes that the innovating firms both create new markets, and suppl these markets initially from a domestic and then from a foreign locatio and, in so doing, they may induce a certain response from other firms an create a market structure which may influence future locational decision Here a distinction between *leading* and *following* firms is necessary (Kindle berger, 1969), as the market size and structure are both dynamic concepts

In a price-taking competitive situation, all profit-maximizing firms wi aim to produce an output at which marginal cost equals price. To do this they may require to produce in one or more locations, depending on th relationships between production costs as output increases and transpor costs as distance increases There are no leaders or followers In ar imperfect market, the firm can influence the character of its market, an hence its optimal location As far as producing overseas is concerned, th firm may do so to gain an advantage over existing producers, or forestal new competition, or to protect its market share even though the rate o return on *new* investment may be very small. In other words, the choic between exports and foreign production will not be taken on purely cost criteria; consideration will also be given to the effects of local production or the market structure in which the investing firm competes and its ability to sell in an imperfectly competitive situation In stressing this factor, location theory is useful, though both the ability and desire of the ME to gain a foothold in a market may be influenced by the fact that it *is* an ME, and explaining the implications of this falls outside the scope of the theory

Empirical studies on location relevant to MEs have so far fallen into three main groups. First, there are those which have sought to evaluate the importance of specific factors affecting the location of either foreign investment or production of MEs. These include Balassa (1967), Horst (1972), Jud (1973), NICB (1961) (costs), Kreinin (1967*b*) (anti-trust legislation in investing countries), Krause (1972) (economic integration in host countries), Stobaugh (1969) (investment climate), Scaperlanda and Mauer (1969) and Schöllhammer (1972) (size of markets), Caves and Reuber (1971) and Morley (1966) (market growth), McAleese (1972) and Falise and Lepas (1970) (investment incentives), Vernon (1972) (threat of competitive firms), and in a paper published earlier this year (Dunning 1973*a*) I tried to

examine some of the principles underlying Britain's entry into the E E C. on the location of international firms in the enlarged Community It is difficult to generalize from these studies, such econometric work as has been done seems to point to the size and growth of market as the single most important demand variable influencing foreign investment (Parry and Ahlburg, 1973).

The second group of studies have adopted a sectoral approach and looked at factors influencing the location of foreign enterprises in particular countries, e g. Stonehill (1965), Brash (1966) (Australia), Daniels (1972) (United States), Deane (1970) (New Zealand), Forsyth (1972) (Scotland), Safarian (1966) (Canada), Schreiber (1970) (Taiwan), or industries, e g. Boranson (1970) (motor vehicles), Harman (1971) (computers), Hufbauer (1966)(synthetic materials), Stobaugh (1973)(petro-chemicals), Tilton (1973) (semi-conductors) and Wortzel (1973) (pharmaceuticals), though most of these, as we have seen, have tended to be an extension of the survey approach. A third group of economists has been interested in location of industry as a feature of international competitiveness (Hirsch, 1967, 1973, Clark, Wilson, and Bradley, 1969, Dunning, 1971, 1973*a*) and, in so doing, have given some attention to the way in which location is influenced by the ownership of firms.

Of the more recent attempts to incorporate the activities of MEs into the general framework of location theory, those of Hymer (1970, 1972*b*), Murray (1973), and Vernon (1973) deserve special mention Vernon argues that the determinants of locational strategy of MEs will vary according to the stage of the product cycle which they are in In both the initial stage of innovative oligopoly and in the final stage of mature oligopoly, their behaviour accords most closely with the interdependence model In the intermediary phases where oligopoly exists with some degree of price competition, cost considerations are likely to be more important As, however, MEs tend to be concentrated in oligopolistic industries and are an important influence on the form of the product cycle, Vernon claims that location theorists should place rather more stress on the interdependence model

A rather different approach is taken by Hymer and Murray who perceive that, parallel to the increasing concentration of firms within industry, there is a trend towards the increasing spatial hierarchy of economic activity. MEs are accelerating this trend: on the one hand routine manufacturing or marketing activities are being dispersed according largely to cost criteria (e g. why do U S firms choose to produce transistor radios and cameras in Taiwan (rather than, e.g , Mexico) for export back to the U S ?), on the other, certain activities, e g top level administration, policy formulation, decision-taking and risk-hedging operations, and the specialized inputs which serve these, e g. technical and financial information, management

expertise, skilled labour, etc., are being increasingly centralized. The spatial interdependence arising from these trends, and particularly the agglomeration of the higher order functions has important implications both for the distribution of income earned by MEs (and their affiliates), and of economic power between nations

Standard location theory is generally concerned with these issues but tends to confine itself to explaining the location of plants of single-product, national firms. Moreover, within its analysis, many of the unique qualities associated with the ME, e g. its ability to shift inputs, such as human capital information and knowledge, across national boundaries at low or zero costs are not brought out Product acceptability is also assumed to be interdependent of the location of supply Because, too, received theory treats the distribution of resources as fixed, it cannot incorporate the situation, often common to MEs in the resource based industries, in which firms can themselves affect this distribution, e g. by their pricing policies and/or exploitation policies This is particularly true of operations of MEs in the less developed countries For these reasons, location theory can give only a partial answer to the question 'why international production?'

5. Industrial organization and market structure

(a) *Concepts and analysis*

The approaches to 'why international production?' so far discussed have been concerned with identifying and evaluating the variables which influence firms in the location of their foreign investment and/or productive activities. The type of answers to the question 'why international production?' they tend to elicit are 'because the prospects for profits or growth are promising' or, focusing more on determinants than goals, 'because foreign labour costs are lower' or 'because there are barriers to exports, etc ' or 'because only by so doing can we protect our competitive position' The following paragraphs attempt to get beyond these indicative variables, and instead of asking 'what causes firms to produce abroad?'—which, in general, can be answered within the existing framework of capital, location, or trade theory—to ask 'under what conditions will particular markets be supplied by the foreign affiliates producing in that market rather than by indigenous firms or imports?'.

In answering this latter type of question it seems to me a complementary approach, viz. that of industrial organization theory, is needed This not only recognizes that international direct investment involves the transmission of a package of capital, knowledge, and entrepreneurship across national boundaries, but that the ownership and control of the organizing unit of this package, i e the foreign affiliate, is domiciled in a different

country (or countries). This immediately suggests distinctiveness on the part of these affiliates *vis-à-vis* indigenous companies

To simplify our analysis, assume there are only two countries (A and B) in the world and that we wish to identify both the *location* and *ownership* of firms manufacturing one particular product—say a drug Assume, too, that there are three ways the market in each country may be supplied, viz by the production of indigenous firms, by imports from firms with production units in the other country, and by the production of affiliates of these firms located in the local market What will determine the extent to which Country A's firms will supply Country B's market from production units located and owned by them in Country B or the extent to which Country B's firms will supply Country A's market by plants located and owned in Country A?

The way in which the question is phrased suggests that there are two primary determinants of the amount of international production The first is the extent of the market in each country and the second is the competitiveness of foreign affiliates *vis-à-vis* indigenous and non-resident firms To simplify matters still further, suppose both the size of the market and the price of the drug are fixed and identical in both countries We are left, then, with deciding how the market is shared between the three groups of firms

Take, first, some extreme situations Suppose transport costs (or other barriers to trade) between Country A and Country B are such that it pays neither country to export to the other This means that each market will be supplied from local production units How will the production be shared between indigenous firms and the affiliates of foreign firms? If it is a question of costs, what will determine whether these favour one group of firms or the other?

Again, take an extreme case Assume that the production of the drug requires knowledge of a formula which is the sole property of firms in Country A. In the absence of licensing or similar arrangements, Country A's firms will supply the market in both Country A and Country B until indigenous producers in Country B are both willing and able to innovate and produce a substitute product. If they do not, then the market will continue to be supplied by Country A's firms

In this particular example, international production (by Country A's firms in Country B) arises because of two absolute barriers, (*a*) the export of goods from Country A to B, (*b*) the inability of indigenous firms in Country B to produce a competitive product.

Now examine the opposite extreme Suppose transport costs are zero and that there are no barriers to production facing firms in either country. Since knowledge is freely and instantaneously transferable, production

functions will be the same in both countries. In this situation, input p will determine relative costs. Suppose these strongly favour Countr Then in a perfectly competitive situation, it could well be that all proc tion will be concentrated in Country B and that Country B's firms supply Country A's market through exports. Since production is zero Country A, it is unlikely there will be any firms in Country A who wo wish to invest in Country B, because of the additional risks and costs operating in a different political and economic environment at a distan from its decision-taking centre (Kindleberger, 1969) It is still possible firms in Country A to invest in Country B's firms, but the investme would be a portfolio kind

In between these two extreme cases, there are a host of intermedia situations, each of which will reflect a combination of the ease or difficu of supplying a particular market with a product (or group of produc from alternative locations, and the ease or difficulty with which firms different ownership can supply the product (or group of products) from t same location

Most of the research on international direct investment has been co cerned with explaining the second characteristic in terms of monopolis competitive theory (i e firms of different nationality but producing in t same location) They are succinctly summarized by Kindleberger (196 Caves (1971), and Gray (1972) Expressed in terms of net advantages M or their affiliates possess over indigenous firms, the former are usua considered under four headings

(i) an easier or cheaper access to knowledge and information,
(ii) an easier or cheaper access to factor inputs,
(iii) a better access to markets or to the saleable characteristics products e g brand names,
(iv) economies of scale and vertical integration

It will be observed, however, that some of these advantages may be e joyed by all branch plants, irrespective of the nationality of the investi firm, and are to do with the internal and/or external economies of si (Coase, 1937; Penrose, 1958) Others arise because the affiliate is part a *foreign* company and a third group because the affiliate is part of an int grated multinational complex of operations

These advantages, according to the Gray Report on *Foreign Dir Investment in Canada*, confer on the ME (or its affiliates) an element distinctiveness which gives them an edge over their competitors (or pote tial competitors) *in similar locations* Essentially, they are enterpris specific, i e they are not transferable between firms, and are a function their character and ownership. The report perceives that some firms a

countries tend to possess the type of qualities which spawn distinctive advantages more than others. These include firms in research-intensive industries and those producing differentiated products, while countries with large markets, a competitive environment, a rapid rate of technological innovation, etc, also tend to be more distinctive Yet small countries may possess distinctive advantages in particular industries or spawn firms with distinctive advantages *within* industries, which explains why trade and investment can be multilateral

Other economists have also probed into the unique characteristics of MEs Johnson (1968), for example, has stressed the importance of the role played by enterprise-specific knowledge; Caves (1971, 1973) adds to this product differentiation, equally important may be the advantages of multinationalism in terms of economies external to the particular production unit but internal to the ME, particularly in *industries* which are research intensive and afford greatest scope for integration, and in firms which are globally integrated in their operations, including those arising from integration H Peter Gray (1972) elaborates on this point, and distinguishes between 'aggressive' and 'defensive' motives for investing abroad The former seek to increase the economic rent of the investing firm largely by the means just described, and/or by obtaining a higher rate of return on capital than is available domestically Defensive investments are those which are made to protect some level of profit (or growth rate) attained earlier (Gray, p 77) These include investments undertaken to preserve a foreign market previously served by exports, and to aquire a safe source of raw materials Finally, some of the younger economists, e g Wolfe (1971), Horst (1972*b*, 1973), Parry (1973), Knickerbocker (1973), extending the earlier work of Edith Penrose (1958, 1968), are interesting themselves in the factors influencing the growth of the multinational firm, the form it takes, and the importance of market structure in influencing both

Rather less attention has been paid to the way in which firms exploit their distinctive advantages Where production remains in the hands of the firm with the advantages, this comes back to a question of location theory, but, in some cases, it may be preferable to sub-contract or to license foreign producers, or to engage in some other scheme to maximize the economic rent on distinctiveness (Hymer, 1972*b*). Though the literature (Kolde *et al*, 1968) is full of examples of the conditions under which licensing is likely to be preferred to direct investment as a means of exploiting foreign markets, there has been little systematic attempt to formalize these into the theory of marketing.

(*b*) *The Aliber thesis*

Before considering some empirical work on the industrial structure

approach, I would like to refer briefly to the currency area approach to international production, which is reflected chiefly in the writings of Aliber (1970, 1972). I can deal with it briefly not because I do not think it important, but because I think it can be accommodated into the conceptual scheme of the industrial structure approach it is also of direct relevance to capital theory

Of all approaches, Aliber is the one which recognizes some of the special aspects of foreign investment which are absent from domestic investment Of these, investment in a different currency area is the most obvious. As I understand it, Aliber is not concerned with explaining investment in the *same* currency area, e g. U.K. investment in Australia, but only where the assets are in different currencies. Since the value of any one currency fluctuates over time it immediately follows that in addition to the variables which influence the worthwhileness of an investment in the local currency, its value in relation to other currencies has to be considered. A rate of return of 10 per cent with a currency that devalues by 5 per cent is worth 5 per cent less the depreciated value of the assets in other currencies. When they invest, firms will then capitalize their income streams taking account of these uncertainties. They will also affect the worthwhileness of trade relative to investment (though observe that a devaluing currency will also affect the worthwhileness of trade as well), but fundamentally, the relationship with their competitors. In other words, the Aliber approach adds to the theory of industrial organization, but I do not think it supplants it [1]

Surveys have revealed that the value and expected variability of the exchange rate are taken account of by firms in deciding the location of their investment, neither is there any doubt that short-term movements of resources are strongly influenced by monetary considerations International companies have been seen to be both at an advantage and a disadvantage in this respect (Manser, 1973)

(*c*) *Recent empirical work*

As yet there have been few attempts to test systematically the type of hypothesis which the above approach suggests There has been a good deal of descriptive analysis and casual empiricism, mainly contained in case studies of countries and industries, and some hints from related studies on technology (Gruber, Mehta, and Vernon, 1967, Mansfield, 1973) Primarily, the lack of progress is due to data deficiencies, but also the subject has generally lacked appeal to economists interested in market structure and

[1] Important to these discussions is *the numéraire* in which the ME keeps its accounts Moreover, one needs to distinguish between the change in parities due to shifts in the terms of trade needed for balanced payments, and those due to differential rates of inflation I am indebted to H. Peter Gray for reminding me of this distinction.

location theory However, recent contributions by Parry (1973) on the determinants of foreign investment in Australian industry, and by Horst (1972a) in which he produces a model to explain the ways in which U S. firms exploit their Canadian and European markets—viz by exports or direct investment—provide useful starting-points

Perhaps the most rewarding attempt to pinpoint the special characteristics of MEs has been that of James Vaupel (1971) in an examination of the 491 largest U S companies [1] Vaupel classifies these companies into three groups, viz national enterprises (NEs) i e those which manufacture only in the U S , transnational enterprises (TNEs)—i e those which manufacture in at least one foreign country but in fewer than six and multinational enterprises (MEs)- i e those which manufacture in at least six foreign countries For the year 1964, there were 125 NEs, 194 TNEs, and 172 MEs He found that MEs had certain distinctive characteristics, for example, they funded 2·4 per cent of their sales on research and development (compared with 1 6 per cent for TNEs and 0 6 per cent for NEs), they spent 2 5 per cent on advertising (compared with 1·9 per cent for TNEs and 1·7 for NEs), they earned net profits of 8·9 per cent on invested capital for the period 1960/4 (compared with 7·3 per cent for TNEs and 6 7 per cent for NEs), their average sales were $460m (compared with $200m for TNEs and $160m for NEs), they were more diversified in their product structure,[2] they recorded a higher exports/sales ratio (6 4 per cent, cf 5 5 per cent) and they paid higher annual wages in the U S ($6,841, cf $6.774)

From the angle of recipient countries, a number of studies have examined the comparative behaviour of foreign affiliates and indigenous domestic firms in Denmark, Holland, and Israel and found that the former were larger, more capital- and skill-intensive and exported a higher proportion of their total output By contrast, Cohen's study of foreign and local firms in South Korea, Taiwan, and Singapore (Cohen, 1973) showed that while the foreign affiliates exported more, they also imported more, and had a lower net output per head Other studies of foreign affiliates in developing countries reveal there is no clear pattern to their capital/labour ratios (*vis-à-vis* indigenous firms) (Strassman, 1968; Pack, 1972; and Wells, 1972b) or to their record as wage payers (Katz, 1972) In our own most recent research on U S investment in the U K (Dunning, 1973b), we attempted to analyse and explain the industrial distribution of the 500, largest U S manufacturing affiliates Tables II and III present details of the distribution of sales of U S affiliates in forty sectors of U K industry in 1970/1, their concentration coefficients, and certain supply and marketing

[1] As listed by *Fortune*.

[2] The measure chosen here was the number of 2, 3, and 5 digit industries in which they operated the results were MEs 5, 10, and 22, TNEs 4, 7, and 15, and NEs 2, 3, and 8

TABLE II

Supply characteristics of industrial distribution of U S. affiliates

	% of total sales of U S affiliates[1]	U S. sales concentration coefficient[2]	1 £	2 %	3 %	4 %	5 %
Food, drink, and tobacco	16 9	0 90	234 8	21 5	0 24	1·12	2·12
Food	8 5	0 77	196 3	19 9	nas	1 00	2 40
Drink	0 3	0 07	396 8	28 6	nas	1·29	2 07
Tobacco	8 1	2 38	249 2	22 5	nas	0 92	1 53
Chemicals	18 6	1 69	666 1	38 7	1 98	1 22	2·52
Mineral oil refining	6 8	3 09	2,313 9	32 0	1 37	1 03	1·03
General chemicals (including dyestuffs and pigments)	1 9	0 58	1,084 4	38 2	nas	1 02	0·79
Pharmaceutical chemicals and preparations	2 7	3 00	445 5	43 9	} 3 54	1 24	7 96
Toilet preparations	0 6	2 00	172 5	40 6		1 02	16 39
Soap and detergents	1 5	2 50	377 8	45 1	nas	1·10	6 65
Synthetic resins and plastics	2 9	2 64	789 1	37 4	2 73	1 03	1·14
Other chemicals	2 2	0 88	226 5	36 9	nas	0 99	1 37
Metal manufacture	4 3	0 46	300 5	23 5	0 25	1 10	0 17
Non-electrical engineering	16 0	1 57	128 4	31 9	1 33	nas	0 63
Agricultural machinery	0 7	3 50	174 3	42 9	1·14	1 12	nas
Machine tools	1 0	1 43	104 9	31 2	0 57	1 02	nas
Pumps, valves, compressors	0 7	0 88	153 6[3]	35 5	na	1·00	nas
Construction and earth-moving equipment	2 9	4 83	218 8	35 1	} 0 67	0 77	nas
Medical handling equipment	0 5	0 71	88 5	34 5		1 08	nas
Office machinery	2 1	7 00	149 4	32 9	nas	0 98	nas
Other machinery	3 7	1 19	137 4[3]	32 9	nas	1 01	nas
Industrial (including process) plant and steel work	3 0	1·58	82 2	34 6	1 05	1 14	nas
Other non-electrical engineering	1 0	0 59	170 2	25 2	nas	1 00	nas
Instrument engineering	4 8	3 43	105 3	36 1	3 03	nas	1 70
Photographic and document copying equipment	1 5	7 50	169 2	38 9	nas	} 1 10	nas
Scientific and industrial instruments and systems	3 2	3 56	116 2	41 0	nas		nas
Other instrument engineering	0 1	0 33	180 5	21 1	nas	0 92	nas

Electrical engineering	8 5	1 09	117 3	33 0	5·23	nas	1 48
Electrical machinery		0 02	71 2	34 9	2 81	0 94	0 75
Electronic computers	2 1	5 25	381 5[3]	50 9		na	
Other electronic apparatus (inc. telecommunications equipment	4 2	1 35	136 6[3]	35 4	13 3	1 00	1 12
Domestic electrical appliances	1 5	2 14	113 2	30 6	0 93	0 99	5 07
Other electrical goods	0·7	0 30	150 0	25 2	1 16	1 15	1 27
Vehicles	21·3	2 09	185 2	29 6	na	1 16	0 48
Motor vehicle manufacturing	21 2	2 90	249 1	22 5	1 58	1 28	0 53
Other products	0·1	0 03	98 9	40 9	nas	1 00	0 35
Metal goods not elsewhere specified	1 8	0 32	132 4	23 1	0 31	1 09	0 64
Textiles and clothing	1 5	0 17	113 0	15 5	0 32	1 20	0 69
Man-made fibres	0 9	1 13	597 4	23 1	nas	1 16	1 47
Other products	0 6	0 08	92 5	15 2	nas	1 06	0 63
Bricks, pottery, glass, cement, etc.	1·1	0 39	243 6	21 9	0 85	1 07	0 71
Abrasives	0 4	4 00	191 0	35 2	nas	1 17	0 91
Other goods	0 7	0 26	245 2	21 4	nas	1 08	0 57
Paper, printing and publishing	1 6	0 25	173 0	29 8	0 23	1 29	1 02
Paper and board	1 1	0 35	251 2	23 0	nas	1 25	0 63
Printing and publishing	0 5	0 15	120 9	34 4	nas	1 32	1 36
Other manufacturing industries	3 7	0 47	120 2	21 9	0 53	1 16	1 10
Rubber	2·4	1 71	211 6	25 9	1 00	1 20	1 27
Other products	1 3	0 20	106 3	21 2	0 35	1 13	1 05
Total manufacturing	100%		191 0	26 2	1 08[4]	1 20	1 18

[1] From EAG survey. [2] Reference in text [3] 1970 only [4] Excludes aerospace

MEASURES

1. Total net capital expenditure per employee (£), 1963 and 1970 average Source *Census of Production 1963*, and *Provisional Results of Census of Production 1970*
2. Proportion of non-operatives (administrative, technical, and clerical employees) to all workers, 1970 (%). Source *Provisional Results of Census of Production 1970*
3. R and D expenditure as a % of sales 1967/9 Source *Annual Abstract of Statistics Census of Production*
4. The labour productivity of the largest 10 per cent of establishments in 1963 divided by the labour productivity of the other 90 per cent as an index of the economies of large scale production Source *Census of Production 1963*
5. The proportion of advertising costs to total sales in 1963 as an index of product differentiation Source *Census of Production 1963*

na = not available nas = not available separately.

TABLE III

Marketing characteristics of industrial distribution of U.S affiliates

	% of total sales of U S. affiliates	*U.S sales concentration coefficient*	1	2	3
Food, drink and tobacco	16.9	0.90	0.95	0.27	85.72
Food	8.5	0.77	1.02	0.12	80.25
Drink	0.3	0.07	1.00	1.89	69.25
Tobacco	8.1	2.38	0.75	0.26	99.52
Chemicals	18.6	1.69	1.03	0.86	77.82
Mineral oil refining	6.8	3.09	1.25	0.34	99.55
General chemicals (including dyestuffs and pigments)	1.9	0.58	0.86	1.19	71.38
Pharmaceutical chemicals and preparations	2.7	3.00	1.38	} 6.08	29.20
Toilet preparations	0.6	2.00	1.52		46.72
Soap and detergents	1.5	2.50	1.02	nas	83.34
Synthetic resins and plastics	2.9	2.64	1.71	1.34	72.61
Other chemicals	2.2	0.88	0.85	0.55	62.58
Metal manufacture	4.3	0.46	0.88	0.81	69.89
Non-electrical engineering	16.0	1.57	1.09	2.01	55.77
Agricultural machinery	0.7	3.50	0.94	2.02	45.90
Machine tools	1.0	1.43	1.20	0.80	25.00
Pumps, valves, compressors	0.7	0.88	nas	nas	22.00
Construction and earth-moving equipment	2.9	4.83	1.81	2.22	57.69
Mechanical handling equipment	0.5	0.71	1.61	1.52	48.87
Office machinery	2.1	7.00	1.49	1.52	na
Other machinery	3.7	1.19	nas	2.33	56.59
Industrial (including process) plant and steel work	3.0	1.58	nas	3.55	50.99
Other non-electrical engineering	1.0	0.59	nas	2.39	78.03
Instrument engineering	4.8	3.43	1.60	1.07	57.34
Photographic and document copying equipment	1.5	7.50	nas	nas	} 49.17
Scientific and industrial instruments and systems	3.2	3.56	nas	nas	
Other instrument engineering	0.1	0.33	nas	nas	80.81

Electrical engineering	8 5	1 09	1 33	1 55	71 84
Electrical machinery		0 02	0 57	3 72	55 20
Electronic computers	2 1	5 25	1 82	1 04	81 90[1]
Other electronic apparatus (inc. telecommunications equipment)	4 2	1 35			77 31
Domestic electrical appliances	1 5	2 14	1 39	1 77	88 26
Other electrical goods	0 7	0 30	1 37	2 35	73 19
Vehicles	21 3	2 09	0 99	3 93	89 60
Motor vehicle manufacturing	21 2	2 90	1 18	7 42	88 52
Other products	0 1	0 03	0 72	1 70	95 23
Metal goods not elsewhere specified	1 8	0 32	1 08	1 21	66 02
Textiles and clothing	1 5	0 17	0 71	0 84	50 47
Man-made fibres	0 9	1 13	1 51	2 03	100 00
Other products	0 6	0 08	0 67	0 80	38 40
Bricks, pottery, glass, cement, etc	1 1	0 39	1 04	2 13	69 20
Abrasives	0 4	4 00	1 17	1 49	65 53
Other goods	0 7	0 26	1 04	2 33	70 42
Paper, printing and publishing	1 6	0 25	1 10	0 34	54 55
Paper and board	1 1	0 35	1 11	0 18	55 98
Printing and publishing	0 5	0 15	1 09	1 79	33 90
Other manufacturing industries	3 7	0 47	1 00	0 56	65 45
Rubber	2 4	1 71	1 10	2 79	89 61
Other products	1 3	0 20	0 98	0 46	50 89
Total manufacturing	100%		0 99	0 99	

[1] 4 firm ratio for 1968

MEASURES

1 The growth of output between 1958 and 1970 divided by the growth of GNP, as an index of the expenditure elasticity of demand Source *Censuses of Production, National Income and Expenditure*

2 Export/import ratio in 1967 as an index of the comparative trading advantage of the U K Source Special tabulations prepared by D T I

3 Five firm concentration ratios in 1963, which illustrate the type of market structure in which U S firms operate Source *Census of Production 1963*

characteristics of the industries in question The concentration coeffici is derived by calculating the percentage of sales of all U.S affiliates accou ted for by a particular industry divided by the percentage of sales of U K firms accounted for by that industry. A concentration coefficient more than one shows that U S affiliates are rather more concentrated that industry than for all industry, a concentration coefficient of less th one suggests the reverse

The five supply features examined in Table II are

1 total net capital expenditure per employee (an average of 1963 a 1970 figures), as an index of the use made of non-human capital, plant and equipment, etc ;
2. the proportion of non-operative to all workers in 1970 as an index the use made of human skills,
3 the value of research and development expenditure (annual avera 1967/9 as a percentage of sales (1968) as an index of technologic intensity,
4 the labour productivity of the largest 10 per cent of establishments 1963 divided by the labour productivity of the other 90 per cent an index of the extent to which large firms enjoy economies of scal
5 the proportion of advertising costs to total sales in 1963 as an ind of product differentiation

The three marketing features examined in Table III are

1 the growth of output between 1958 and 1970 divided by the growt in GNP as an index of the expenditure elasticity of demand,
2 the export/import ratio in 1967 as an index of the comparative tradu advantage of the U K , and
3 the output of the five largest firms in an industry as a proportion the total output of that industry in 1963, which illustrates the ty of market structure in which U S affiliates operate

Tables IV and V summarize the data contained in these tables Table I compares the industrial distribution of U S affiliates with that of U K firms as a whole The value of each characteristic presented in Tables I and III is weighted by the distribution of, first, U K firms, and secon U S affiliates, and then averaged to give the figure set out in Table IV

The conclusions of this exercise are self-evident U.S affiliates tend t be more concentrated in faster-growing and export-oriented industries They are also attracted to the technologically advanced industries, and t those where both capital and advertising expenditure is slightly abov average these are also the industries in which the barriers to entry facin indigenous firms are likely to be higher than those facing U S affiliates There is, however, no evidence to suggest that their share of industries whic

TABLE IV

Summary of characteristics of all U K firms and U S affiliates

	Average figures	
	U K firms	U S. affiliates
Supply characteristics		
1. Net capital expenditure per employee	£191 0	£221 2[1]
2. Non operatives/total workers	26 2%	30 3%[1]
3. R and D expenditure as a % of sales	1 08	1 60[1]
4 Economies of scale	1 09[2]	1 09[2]
5 Advertising expenditure as a % of sales	1 18	1 33[1]
Marketing characteristics		
1. Output growth/GNP growth	1 02[2]	1 14[2]
2 Exports/Imports ratio	1 23[2]	1 66[2]
3 Concentration ratio	70 9[2]	74 7[2]

[1] Values of characteristics from Table II weighted by distribution of U S sales/employment

[2] Values of characteristics from Tables II and III weighted by distribution of U K and U S. sales respectively.

SOURCE Tables II and III

TABLE V

Classification of supply and marketing characteristics of U S affiliates by concentration coefficient

	Supply characteristics[1]					Marketing characteristics[2]		
	1 £	2 %	3 %	4	5	1	2	3
Group 1 (10 industries) U S sales concentration coefficient 7 50 to 2 90	440 0	37 5	2 75[3] (Groups 1 and 2)	1 00[4]	2 42[5]	1 35[4]	2 30[7]	54 41[4]
Group 2 (11 industries) U S sales concentration coefficient 2 64 to 1 13	270 0	32 7		1 05	4 27[4]	1 28[4]	1 90[4]	71 81
Group 3 (9 industries) U S sales concentration coefficient 0 88 to 0 33	291 4	28 6	0 71[4] (Groups 3 and 4)	1 04	1 07[5]	1 06[4]	0 97[7]	63 30
Group 4 (9 industries) U S sales concentration coefficient 0 32 to 0 02	157 1	27 2		1 12	0 96	0 95[4]	1 80	61 37

[1] For definitions of supply characteristics see Table II

[2] For definitions of marketing characteristics see Table III

[3] 11 industries only

[4] 8 industries only

[5] 5 industries only

[6] 6 industries only

[7] 7 industries only

SOURCE Tables II and III

benefit from the economies of scale is greater than that of U K companies, and their market structure is only slightly more oligopolistic

Table V classifies these same characteristics by four groups of U S affiliates Group 1 consists of the ten affiliates with the highest concentration ratios (from 7·50 to 2·90), Group 2 of the eleven affiliates with the next highest concentration ratios (from 2 38 to 1·19); and Groups 3 and 4 of the

eighteen firms with concentration ratios of below 1. The results of this exercise confirm the general pattern already stated.

What, next, of an explanation for the structure of U S. participation in U K. industry? Two propositions might be tested First, that U S firms will produce most in the U.K in those industries where both the growth and/or profit potential is favourable relative to that of exploiting foreign markets by other means, e g exports The second is that U S firms will invest in those industries where the comparative advantage of the U S firms is greatest *vis-à-vis* that of U.K firms

While data limitations preclude any systematic testing of these hypotheses, certain pointers may be obtained by looking again at some of the statistics contained in Tables I, II, and III and also some additional figures set out in Table V

Index (1), for example, expresses the sales of U S affiliates in the U K as a ratio of U K imports from the U S This shows very clearly that this ratio is highest in those sectors where the U S concentration coefficient is the highest Index (2) presents details of the U K. nominal tariff on the imports of various goods, there appears to be no obvious relationship between the size of the tariff and either the U S concentration coefficient or the previous index (An exercise by Horst (1972) which used estimates of effective rates of protection came to broadly similar conclusions) Index (3) gives details of the total productivity of U S affiliates and suggests that the affiliates do tend to concentrate where this is highest, and the remaining three indices (4) to (6) present data which are intended to be surrogates for barriers to entry into particular industries Here, the proposition is that these are likely to be the greatest in those industries where the content of productive knowledge is important, or where the costs of entry are high, or where product differentiation is most marked. The data in Table IV which summarize our conclusions, lend some corroboration to this hypothesis

The data analysed hints of some *raison d'être* to the structure of U S participation in U K. industry, but it does little more than this. There are various reasons for this but perhaps the main ones are (i) that the industrial classification is not fine enough for us to be able to say much about the relationship between investment and exports as a means of exploiting a market, (ii) other locational variables, noticeably transport and labour costs, are ignored, and (iii) sales are not always a good guide to the value added by the firms

The first problem is particularly acute where firms are multi-product and investment and exports may complement as well as substitute for each other. This suggests that international production gives awareness to the products only produced by the investing company in local markets; more-

over, parts and components might be required. The evidence on the relationship between exports and foreign investment at a macro level is inconclusive (Hufbauer and Adler, 1968; Reddaway, 1968), however much at a micro level clearly they may substitute for each other

Such relationships become rather more complex when the activities of MEs become *industrially* and *regionally* integrated Taking the latter point first, a company may replace exports to half a dozen European countries by setting up a plant in one of these and supplying the entire market from there In this case, the production implications for the country in which the plant is located will be much greater than the replacement of imports might suggest, while, in other countries, European imports will replace U S imports

As to *industrial* integration, this will take the pattern mentioned earlier of *horizontal* or *vertical* specialization of products or processes In our earlier example, if the firm owned by Country A manufactured two drugs it might decide to concentrate the production of one in Country A and supply both countries from that plant, and concentrate the production of the other in Country B and supply both markets from there Or it may engage in first-stage production in a plant in Country A, export the semi-processed good to Country B, have it made up there and then sold in both countries In this case there is intra-group trading as well as two-way investment. Seeking to explain the determinants of international production then becomes extremely complex, although basically it is an exercise in the theory of the growth of the firm (Penrose, 1958, Horst, 1973), and, as we have said, the fact that an affiliate of a foreign firm may possess net advantage over local producers may lie in the nature of branch plant economies, and enterprise-specific integration An indigenous competitive firm, for example, might have to engage in setting up costs already incurred elsewhere in the firm's organization

The desire to achieve the economies of industrial or regional integration is, of course, less an explanation of the *initial* decision of an enterprise to set up a foreign production unit as a strategy that an established company might pursue Many American firms already operating in different parts of Western Europe are now rationalizing their production programmes in such a way as we are likely to have important locational repercussions, and will almost certainly increase the volume of intra-group trade between the individual European affiliates But again, here, no new principles of growth are involved

(*d*) *Lines for further research*

One conclusion which follows from the previous paragraphs is that the question 'why international production?' is now less interesting than 'why

Table VI

Indices of comparative advantage of U S affiliates in U K

	% of total sales of U S affiliates	U S sales concentration coefficient	1	2 %	3	4 £	5 %	6 %
Food, drink, and tobacco	16.9	0.90	39.6	na	1.38	234.8	0.24	2.12
Food	8.5	0.77	21.6	na	1.38	196.2	nas	2.40
Drink	0.3	0.07	28.8	na	0.93	396.8	nas	2.07
Tobacco	8.1	2.38	342.4	na	1.41	249.2	nas	1.53
Chemicals	18.6	1.69	9.4	15.6	1.34	666.1	1.98	2·52
Mineral oil refining	6.8	3.09	44.4	na	na	2,313.9	1.37	1.03
General chemicals (including dyestuffs and pigments)	1.9	0.58	2.4	18.5	1.61	1,084.4	nas	0.79
Pharmaceutical chemicals and preparations	2.7	3.00	28.5	15.3	1.39	445.5	3.54 (Pharmaceutical chemicals and Toilet preparations combined)	7.96
Toilet preparations	0.6	2.00	23.2 (Toilet preparations and Soap and detergents combined)	17.3	1.46	172.5		16.38
Soap and detergents	1.5	2.50		12.2	1.61	377.8	nas	6.65
Synthetic resins and plastics	2.9	2.64	5.7	17.7	1.39	789.1	2.73	1.14
Other chemicals	2.2	0.88	6.1	12.7	1.74	226.5	nas	1.37
Metal manufacture	4.3	0.46	3.0	10.4	1.06	300.5	0.25	0.17
Non-electrical engineering	16.0	1.57	3.8	16.0	1.24	128.4	1.33	0.63
Agricultural machinery	0.7	3.50	12.9	14.0	0.74	174.3	1.14	nas
Machine tools	1.0	1.43	5.4	16.8	1.18	104.9	0.57	nas
Pumps, valves, compressors	0.7	0.88	3.1	16.2	1.54	153.6	na	nas
Construction and earth-moving equipment	2.9	4.83	7.4	15.5	1.55	218.8	0.67 (Construction and earth-moving equipment and Mechanical handling equipment combined)	nas
Mechanical handling equipment	0.5	0.71	4.3	15.5	1.05	88.5		nas
Office machinery	2.1	7.00	2.2[1]	15.4	1.35	149.4	nas	nas
Other machinery	3.7	1.19	nas	14.7	1·21	137.4	nas	nas
Industrial (including process) plant and steel work	3.0	1.58	nas	17.5	1.10	82.2	1.05	nas
Other non-electrical engineering	1.0	0.59	nas	18.0	1.20	170.2	nas	nas
Instrument engineering	4.8	3.43	6.2	27.5	1.33	105.3	3.03	1.70
Photographic and document copying equipment	1.5	7.50	nas	22.9	1.86	169.2	nas	nas
Scientific and industrial instruments and systems	3.2	3.56	nas	32.0	1.18	116.2	nas	nas
Other instrument engineering	0.1	0.33	nas	27.5	1.20	180.5	nas	nas

Electrical engineering	8.5	1.09	3.3	17.2	1.48	117.3	5.23	1.48
Electrical machinery	..	0.02	nas	18.9	1.80	71.2	2.81	0.75
Electronic computers	2.1	5.25	..	na	2.35	381.5	13.3	1.12
Other electronic apparatus (inc in telecommunications equipment)	4.2	1.35	nas	20.1	1.19	136.6		
Domestic electrical appliances	1.5	2.14	59.9	14.5	1.36	113.2	0.93	5.07
Other electrical goods	0.7	0.30	nas	16.9	1.53	150.0	1.16	1.27
Vehicles	21.3	2.09	15.2	19.3	1.12	185.2	na	0.48
Motor vehicles manufacturing	21.2	2.90	97.6	21.1	1.12	249.1	1.58	0.53
Other products	0.1	0.03	0.1	18.8	1.04	98.9	nas	0.35
Metal goods not elsewhere specified	1.8	0.32	5.3	18.0	1.26	132.4	0.31	0.64
Textiles and clothing	1.5	0.17	4.0	18.5	1.63	113.0	0.32	0.69
Man-made fibres	0.9	1.13	nas	16.0	1.63	597.4	nas	1.47
Other products	0.6	0.08	nas	20.9	1.62	92.5	nas	0.63
Bricks, pottery, glass, cement, etc	1.1	0.39	6.5	18.0	1.23	243.6	0.85	0.71
Abrasives	0.4	4.00	nas	18.5	1.31	191.0	nas	0.91
Other goods	0.7	0.26	nas	17.5	1.19	245.2	nas	0.57
Paper, printing, and publishing	1.6	0.25	2.7	na	1.10	173.0	0.23	1.02
Paper and board	1.1	0.35	3.0	17.2	1.10	251.2	nas	0.63
Printing and publishing	0.5	0.15	2.2	na	1.11	120.9	nas	1.36
Other manufacturing industries	3.7	0.47	9.7	20.7	1.41	120.2	0.53	1.10
Rubber	2.4	1.71	18.4	23.0	1.34	211.6	1.00	1.27
Other products	1.3	0.20	5.2	18.5	1.56	106.3	0.35	1.05
	100%	..	6.9	..	1.26	191.0	1.08	1.18

[1] Electronic computers including office machinery

INDICES

1 Sales of U S affiliates in U K divided by imports into U K from U S Source E A G Survey (sales), O E C D. Commodity Trade Statistics (imports)

2 Nominal tariff (unweighted average for components of group) derived from S S Hen and H. H Liesner *Britain and the Common Market*, Cambridge University Press 1971

3 Total productivity of U S affiliates

4 As measure 1, Table II 5 As measure 3, Table II 6 As measure 5, Table II

na = not available nas = not available separately

the present rate of growth in international production?' or 'why the particular geographical or industrial pattern of international production?' and that future research should be focused on the dynamics of multinational enterprises and comparative studies On the first point, various explanations might be adduced both of the increasing role of such institutions in the world

TABLE VII

Classification of comparative advantage characteristics of U.S affiliates by sales concentration coefficient

<table>
<tr><th></th><th>1</th><th>2
%</th><th>3</th><th>4
£</th><th>5
%</th><th>6
%</th></tr>
<tr><td>Group 1 (10 industries)
U S. concentration coefficient
7 5 to 2·90</td><td>28·5[1]</td><td>19 3[5]</td><td>1·43[5]</td><td>440 9</td><td rowspan="2">2 75[6]</td><td>2 42[7]</td></tr>
<tr><td>Group 2 (11 industries)
U S concentration coefficient
2 64 to 1 13</td><td>19 3[2]</td><td>16 9</td><td>1 49</td><td>270 0</td><td>4 27[3]</td></tr>
<tr><td>Group 3 (9 industries)
U.S. concentration coefficient
0 88 to 0 33</td><td>6 3[3]</td><td>17 0[3]</td><td>1 32</td><td>294·4</td><td rowspan="2">0·71[3]</td><td>1 07[7]</td></tr>
<tr><td>Group 4 (9 industries)
U S concentration coefficient
0 30 to 0 02</td><td>7 6[4]</td><td>18 5[1]</td><td>1 34</td><td>157·1</td><td>0 96</td></tr>
</table>

[1] 7 industries only.
[2] Excludes tobacco: 6 industries only
[3] 8 industries only
[4] 6 industries only
[5] 8 industries only.
[6] 11 industries only
[7] 5 industries only
SOURCE Table VI

economy and their changing character One of these is simply that they tend to be concentrated in the new industries which are growing faster than the average in the world economy. The second is that MEs seem to be more profitable and grow faster than indigenous firms (Dunning and Pearce, 1971) which enables them to acquire the necessary resources for additional growth The third is that as the firms increase in size and become more established, the chances of competitors breaking into the market are less The fourth is that as they grow, the companies often enhance their competitive advantages, sometimes by tightening up on control of market, sometimes increasing integration and so on

All of these are symptomatic of broad trends in industrial structure One of these is the general increase in industrial concentration within particular countries although not for the world as a whole. The proportion of motor cars, petrol, rubber tyres, pharmaceuticals, etc , produced by (say)

the five largest companies in the world has fallen in recent years—largely due to the resurgence in Japanese and European competition (Rowthorn, 1969). There is nothing inevitable about this trend of growth of MEs Anything which reduces the barriers to competition on which these companies thrive may reduce their share of output The end of a patent could mean that a foreign affiliate is no longer protected from indigenous firms, and loses its competitive edge, this has happend in the U K pharmaceutical industry (Cooper and Culyer, 1973) Or a new product might replace an old one which can be more easily produced by competitive companies, the decline of the share of U S affiliates in the foundation garment industry is an illustration here There is a substantial learning process associated with competition engendered by international companies, the declining share of the main U S affiliate in the razor blade industry is a case in point, though, as often as not, competition comes from other international companies

The second line of research which needs pursuing is a more systematic analysis of the distinctiveness of MEs and alternative forms of market penetration, by country and industry [1] Why is it, for example, that although the U K and U S account for 35 per cent of world exports they are responsible for 70 per cent of the world's investment income in 1968? Why is the broad industrial pattern of the Japanese MEs different from that of their U S and European counterparts? (United Nations, 1971) Why do the sales of foreign affiliates/export ratios of countries differ enormously, being, for example, high for the U S, Switzerland, Sweden, and Holland and low in Japan, France, and Italy?, and of industries within countries, e g of motor vehicles and computers with industrial instruments and cotton textiles? Various possible explanations come to mind One is to do with the structure of a country's comparative advantage Where this is in goods which can be easily tradable or can be easily assimilated abroad the percentage might be less Another may have to do with structure of markets; dispersed markets may make foreign production uneconomical while more concentrated markets would not do so A third is to do with the different organizational patterns of MEs of different nationality (Stopford, 1973, Franko, 1972), and a fourth with the attitudes and policies of both exporting countries to exports relative to outward investment, and importing countries to imports relative to inward investment This, in turn, will be related to balance of payments questions If the dollar is in short supply but the yen is plentiful, then under a fixed exchange rate, tariffs might be placed on dollar goods which might encourage defensive investment, while

[1] The work now being undertaken by Raymond Vernon and his colleagues on European and Japanese MEs should prove particularly illuminating in this respect See also Hellman (1970)

Japanese firms can export freely Methods of restricting capital outflows also vary between countries (Cairncross, 1973).

A fourth reason concerns economic conditions in investing or exporting countries Firms do not usually look overseas for markets if ones nearer home can be satisfied And generally they prefer exports to foreign production. The more profitable the opportunities for growth at home, the less foreign markets will be vigorously pursued I believe the lack of German and Japanese foreign investment for a long time since the Second World War can be largely explained in terms of the rapid internal growth of the two economies, and the fact that the undervaluation of their currencies favoured the exploitation of foreign markets by exports rather than by outward direct investment Now these conditions no longer hold, there are signs that both countries are becoming important foreign investors But the extent to which firms face demand pressures in domestic or foreign markets which can be met without production overseas will influence their levels of foreign activities

Lastly, government policy is vitally important. This may be exerted in various ways, both by direct controls (Herring and Willett, 1972) and affecting the value of the variables which influence decision-taking by firms to invest overseas. This is very relevant to the question 'how much international production?' but can also influence 'why international production?'. There are many obvious examples of government affecting the behaviour of international companies and it seems likely that the role will become even more important in the future.

REFERENCES

AHARONI, Y., *The Foreign Investment Decision*, Harvard University Press, 1966.

—— 'The definition of a multinational corporation', *Quarterly Review of Economics and Business*, Autumn, 1971

ALIBER, R Z, 'A theory of direct investment' in C. P. Kindleberger (ed), *The International Corporation*, M.I.T. Press, 1970.

—— 'The multinational enterprise in a multiple currency world' in J. H Dunning (ed), *The Multinational Enterprise*, Allen and Unwin, 1971.

ANDREWS, M, *American Investment in Irish Industry*, Senior Honours thesis, Harvard University, 1972

BALASSA, B., *Trade Liberalisation Among Industrial Countries*, McGraw-Hill, 1967.

BALDWIN, R, 'International trade in inputs and outputs', *American Economic Review*, vol 60, 1970

BANDERA, V N, and WHITE, J. J, 'US direct investment and domestic markets in Europe', *Economica Internationale*, vol 21, 1968

BARLOW, E. R., and WENDER, I. T., *Foreign investment and taxation*, Prentice Hall, 1955

BASI, R. S., *Determinants of US Direct Investment in Foreign Countries*, Kent University Press, 1966.

BEHRMAN, J., 'Foreign associates and their financing' in R. Mikesell (ed.) *US Private and Government Investment Abroad*, Oregon University Press, 1962.

BEHRMAN, J., *Some Patterns in the Rise of the Multinational Enterprise*, University of Carolina research paper, 1969.

BORTS, G. H. and KOPECKY, K. J. 'Capital movements and economic growth in developed countries' in F. Machlup, W. Salant and L. Tarshis (eds), *International Mobility and Movement of Capital*. National Bureau of Economic Research, 1972

BRANSON, W. H., 'Monetary policy and the new view of international capital movements', *Brookings Papers on Economic Activity*, No 2, 1970.

—— and HILL, R. D., *Capital Movements in the OECD Area an Economic Analysis*, O.E.C.D., 1971

BRASH, D., *American Investment in Australian Industry*, Australian National University Press, 1966

BROOKE, M. Z., and REMMERS, H. L., *The Strategy of Multinational Enterprise*, Longmans, 1970.

BRUCK, N. A, and LEES, F A., 'Foreign content of US corporate activities, *Financial Analysts Journal*, vol. 22, Sept /Oct 1966

BRUNO, M, 'Development policy and dynamic comparative advantage' in R Vernon (ed), *The Technology Factor in International Trade*, Columbia University Press, 1970.

CAIRNCROSS, A K, *Control over International Capital Movements*, The Brookings Institution, 1973

CAVES, R, 'International corporations the industrial economics of foreign investment', *Economica*, 1971, vol. 38

—— 'The multinational enterprise and industrial structure' in J H Dunning (ed), *Economic Analysis and the Multinational Enterprise*, Allen and Unwin (forthcoming).

—— and REUBER, G. L. *Capital Transfers & Economic Policy, Canada, 1957/62*, Harvard University Press, 1971

CLARK, C., WILSON, F, BRADLEY, J., 'Industrial location and economic potential in Western Europe', *Regional Studies*, vol 3, 1969

COASE, R H, 'The nature of the firm', *Economica*, NS vol 4, 1937

COHEN, BENJAMIN I, Foreign investment by US corporations as a way of reducing risk', Economic Growth Centre Discussion Paper No 151, Yale University, 1972.

—— 'The role of the multinational in the exports of manufactures from developing countries', Economic Growth Centre Discussion Paper No 177, Yale University, 1973

COOPER, M., and CULYER, A, *The Pharmaceutical Industry*, EAG/Dun and Bradstreet Industry Profile No. 2, 1973.

DANIELS, J. D, *Recent Foreign Direct Manufacturing Investment in the United States*, Praeger, 1972

D'ARGE, R, 'Notes on customs unions and foreign direct investment', *Economic Journal*, vol. 74, 1969.

DEANE, R. S, *Foreign Investment in New Zealand manufacturing* Sweet and Maxwell, 1970

DUNNING, J. H., *Studies in international investment*, Allen and Unwin, Chapters 3 and 5, 1970.

—— *The Multinational Enterprise*, Allen and Unwin, 1971

—— *The Location of International Firms in an Enlarged EEC* Manchester Statistical Society, 1973.

—— *United States Industry in Britain*, an E A G Business Research Study Financial Times, 1973.

—— and PEARCE, R. D., 'The world's largest companies a statistical profile', *Business Ratios*, vol. 3, 1971.

FALISE, M, and LEPAS, A., 'Les Motivations de localisation des investissements internationaux dans l'Europe du Nord-Ouest', *Revue Économique*, No. 1, 1970

FLOYD, J. E., 'International capital movements and monetary equilibrium', *American Economic Review*, vol. 59, 1969.

FORSYTH, D. J. C., *US Investment in Scotland*, Praeger Special Studies in International Economics and Development, 1972.

FRANKO, L. G., *Organisational Change in European Enterprise*, Centre for Education and International Management, Mar. 1972.

Fundación de Investigaciones Económicas Latin Americas (FIEL), 'Las inversiones extranjeras en la Argentina' FIEL, 1971.

GOLDBERG, M. A., 'The determinants of US direct investment in the EEC comment', *American Economic Review*, vol. 62, 1972.

GIERSCH, H., Economic union between nations and the location of industries', *Review of Economic Studies*, vol. xvii(2), 1950

GRAY, H P., *The Economics of Business Investment Abroad*, Macmillan, 1972.

GREENHUT, M, 'The size and shape of the market area of the firm', *Southern Economic Journal*, July 1952.

GRUBEL, H., 'Internationally diversified portfolios: welfare gains and capital flows', *American Economic Review*, vol. 58, 1968

GRUBER, W, MEHTA, D., and VERNON, R., 'The research and development factor in investment of US industries', *Journal of Political Economy*, vol 75, 1967.

HAKAM, A N, 'The motivation to invest and the locational pattern of foreign private industrial investments in Nigeria', *Nigerian Journal of Economic and Social Studies*, vol. 8, Mar. 1966

HARMAN, A. J, 'The international computer industry. innovation and comparative advantage', *Harvard University Press*, 1971.

HECKERMAN, D. G, *The Exchange Risks of Foreign Operations*, Graduate School of Chicago, (mimeo), 1969.

HELLMAN, R, *The Challenge to US Dominance of the International Corporation*, Dunellen, 1970

HERRING, R., and WILLETT, T. D, 'The capital control program and US investment activity abroad', *Southern Economic Journal*, July 1972

—— —— 'The relationship between US direct investment at home and abroad', *Rivista Internazionale di Scienze Economiche e Commerciali*, Anno XX, 1973.

HIRSCH, S, *Location of Industry and International Competitiveness*, Oxford University Press, 1967

—— 'Multinational corporations · how different are they?' in G Bertin (ed), *The Growth of the Large Multinational Enterprise*, Rennes, 1973.

HOGAN, W., *Foreign Investments and Capital Inflows*, The English, Scottish & Australian Bank Ltd, Research Lecture, 1968.

HORST, T, 'The industrial composition of US exports and subsidiary sales to the Canadian market', *American Economic Review*, vol 62, 1972

—— 'Firm and industry determinants of the decision to invest abroad an empirical study', *Review of Economics and Statistics*, vol 14, 1972.

—— 'The multinational enterprise and the theory of the firm' in J. H. Dunning (ed), *Economic Analysis and the Multinational Enterprise*, Allen and Unwin (forthcoming).

HUFBAUER, G. C., *Synthetic Materials and the Theory of International Trade*, Duckworth, 1966

—— and ADLER, M, *Overseas Manufacturing Investment and the Balance of Payments* US Treasury Department, 1968.

—— 'The impact of national characteristics and technology on the commodity composition of trade in manufactured goods' in R. Vernon (ed), *The Technology Factor in International Trade*, Columbia University Press, 1971.

HYMER, S., *The International Operations of National Firms a Study in Direct Investment*. Unpublished doctoral dissertation, M I.T, 1960.

—— 'The multinational corporation and the law of uneven development' in J. N Bhagwati (ed.), *Economics and the World Order*, World Law Fund, 1970.

—— 'The internationalization of capital', *Journal of Economic Issues*, Mar. 1972.

—— 'United States investment abroad' in Peter Drysdale (ed.), 'Direct foreign investment in Asia and the Pacific', *Australian National Pacific*, 1972

JOHNS, B. L., 'Private overseas investment in Australia profitability and motivation', *Economic Record*, vol. 43, 1967.

JOHNSON, H G., 'International capital movements and economic policy' in *Essays in Honour of Marco Fanno*, Padova, Italy, 1966

—— *Comparative Cost and Commercial Policy for a Developing World Economy*, The Wicksell Lectures, 1968.

—— 'The efficiency and welfare implications of the international corporation' in C. Kindleberger (ed), *The International Corporation*, M I T Press, 1970.

JORGENSON, D U., 'Capital theory and investment behaviour', *American Economic Review*, vol 53, 1963.

—— and SIEBERT, C. D., 'A comparison of alternative theories of corporate investment behaviour', *American Economic Review*, vol 58, 1968

JUD, C D, 'An empirical study of the industrial composition of US exports and foreign subsidiary sales', paper read to South Western Economic Association, March, 1973

KATZ, J, 'Importacion de tecnologia, aprendizajo local e industrializacion dependiante', *Buenos Aires Instituto Di Tella*, 1972

KEESING, D, 'Labour skills and comparative advantage', *American Economic Review*, vol. 56, 1966

KENEN, P B., 'Short term capital movements and the US balance of payments' in *The United States Balance of Payments*. Hearings before the Joint Economic Committee, 88 Congress, First Session, 1963

KINDLEBERGER, C, *American Business Abroad*, Yale University Press, 1969

KLEIN, R. W, 'A dynamic theory of comparative advantage', unpublished manuscript forthcoming in *American Economic Review*

KNICKERBOCKER, F. T, *Oligopolistic Reaction and the Multinational Enterprise*, Harvard University Press, 1973

KOLDE, E, *International Business Enterprise*, Prentice Hall, 1968.

KOPITS, G., 'Dividend remittance behaviour within the international firm. a cross country analysis', *Review of Economics and Statistics*, Aug 1972

KRAUSE, L R, *European Economic Integration and the US*, Washington, 1972

KREININ, M E, 'Freedom of trade and capital movement some empirical evidence', *Economic Journal*, vol. 75, 1965.

—— 'Trade arrangements among industrial countries effects on the United States' in B Balassa (ed), *Studies in Trade Liberalisation*, Johns Hopkins Press, 1967

KREININ, M, *Alternative Commercial Policies Their Effect on the American Economy*, Michigan State University Press, 1967.

KWACK, S. Y, 'A model of US direct investment abroad A new classical approach', *Western Economic Journal*, forthcoming

LEARNER, E. E., and STERN, R. M., 'Problems in the theory and empirical estimation of international capital movements' in F. Machlup, W Salant, and L Tarshis (eds), *International Mobility and Movement of Capital*, National Bureau of Economic Research, 1972.

LEFTWICH, R. B, 'Foreign direct investments in the United States', *Survey of Current Business*, Feb. 1973

LEVY, H., and SARNAT, M., 'International diversification of investment portfolios', *American Economic Review*, vol. 60, 1970.

LINTNER, J., 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets', *Review of Economics and Statistics*, vol 47, 1965

LOSCH, A., *The Economics of Location*, New Haven, Halo University Press, 1954.

MCALEESE, D., *Capital Inflows and Direct Foreign Investment in Ireland 1947 1970* (unpublished paper) 1972.

McGraw Hill Economics Department, *Overseas Operations of US Industrial Companies 1960–1961*, McGraw Hill, 1961.

MANSER, W, *The Financial Role of Multinational Enterprises*, International Chamber of Commerce, 1973

MANSFIELD, E, 'The multinational enterprise and technological change' in J. H Dunning (ed), *Economic Analysis and the Multinational Enterprise*, Allen and Unwin (forthcoming).

MARKOWITZ, H, *Portfolio selection Efficient Diversification of Investment*, Wiley, 1959.

MAY, H K., and ARENA, J. A F., *Impact of Foreign Investment in Mexico*, National Chamber Foundation and Council of the Americas, 1971.

MELLORS, J, *International Tax Differentials and the Location of Overseas Direct Investment a Pilot Study*, University of Reading Research Papers in International Investment and Business, No 4, 1973.

MILLER, R R, and WEIGEL, D. R, 'The motivation for foreign direct investment', *Journal of International Business Studies*, vol 3, Fall, 1972

MILLER, N. C, and WHITMAN, M V. N., 'A mean variance analysis of US long term portfolio foreign investment', *Quarterly Journal of Economics*, vol 84, 1970

MODIGLIANI F, and MILLER, M., 'The cost of capital, corporation finance and the theory of investment, *American Economic Review*, vol 48, June 1958

MOOSE, J, *US direct investment abroad in manufacturing and petroleum a recursive model*, unpublished doctoral thesis, Harvard, 1968

MORLEY, S, *American Corporate Investment Abroad Since 1919*, unpublished doctoral dissertation, University of California (Berkeley), 1966.

MUNDELL, R A, 'International trade and factor mobility', *American Economic Review*, vol 47, 1957

- - - 'The monetary dynamics of international adjustment under fixed and flexible exchange rates', *Quarterly Journal of Economics*, vol 74, May 1960.

MURRAY, R, 'Underdevelopment, international firms and the international division of labour' in *Towards a New World Economy*, Rotterdam University Press, 1973.

National Industrial Conference Board, *Costs and Competition American Experience Abroad*, N I C.B, 1961

OHLIN, B, *Inter-regional and International Trade* (revised edition), Harvard University Press, 1967

PACK, H, *Employment in Kenyan Manufacturing Some Microeconomic Evidence* (mimeo), 1972.

PARKER, J, *The Diffusion of Technology and the Multinational Enterprise* (unpublished University of Exeter Ph D thesis), 1973

PARRY, T., *Technology and Performance of the Foreign Subsidiary in Australia* (paper presented to New Zealand Association for the Advancement of Science, Aug 1972)

—— and AHLBURG, D A, *Determinants of US Direct Investment in Australian Manufacturing Industry* University of Reading Research Paper in International Investment and Business, No. 2, 1973

PENROSE, E T, *The Theory of the Growth of the Firm*, Basil Blackwell, 1958.

—— *The Large International Firm in Developing Countries*, Allen and Unwin, 1968

PERLMUTTER, H, 'The tortuous evolution of the multinational company', *Columbia Journal of World Business*, Jan./Feb 1969.

PIPER, J. R, 'How US firms evaluate foreign investment opportunities', Michigan State University *Business Topics*, Summer 1971.

POLK, J, 'The new world economy', *Columbia Journal of World Business*, Jan /Feb 1968.

—— *World Companies and the New World Economy* (unpublished paper prepared for discussion group at Council for Foreign Relations, New York), 1971

—— MEISTER, I. W, and VEIT, L. A., *US Production Abroad and the Balance of Payments*, N I.C.B, 1966.

POPKIN, J., *Interfirm Differences in Direct Investment Behaviour of US Manufacturers*, unpublished doctoral dissertation, University of Pennsylvania, 1965.

POSNER, M. V., International trade and technical change, *Oxford Economic Papers*, vol 13, 1961

PRACHOWNY, M J, 'Direct investment and the balance of payments of the US a portfolio approach' in F Machlup, W. Salant, and L Tarshis (eds) *International Mobility and Movement of Capital* National Bureau of Economic Research, 1972.

QUINN, D, 'Scientific and technical strategy at the national and major enterprise level', paper for U N E S.C O. symposium on *The Role of Science and Technology in Economic Development*, Paris, 1970

REDDAWAY, W B, POTTER, S. T, and TAYLOR, C T, *The Effects of UK Direct Investment Overseas*, Cambridge University Press, 1967 and 1968

RICHARDSON, J. D., 'Theoretical consideration in the analysis of foreign direct investment', *Western Economic Journal*, Mar. 1971.

—— 'On going abroad, the firms initial foreign investment decision', *Quarterly Journal of Economics and Business*, vol. 11, 1972.

ROBERTSON, D, 'The multinational enterprise trade flows and trade policy' in J H. Dunning (ed), *The Multinational Enterprise*, Allen and Unwin, 1971

ROBINSON, H. J, *The Motivation and flow of Private Foreign Investment*, Stanford Research Institute, California, 1961

ROLFE, S., *The International Corporation*, International Chamber of Commerce, 1969.

ROWTHORN, R., *International Big Business, 1957–1967*, Cambridge University Press, 1969

SAFARIAN, A. E, *Foreign Ownership in Canadian Industry*, McGraw Hill, 1966

SCAPERLANDA, A E, 'The EEC and US foreign investment some empirical evidence', *Economic Journal*, vol 77, 1967

—— and MAUER, L J, 'The determinants of US direct investment in the EEC', *American Economic Review*, vol 59, 1969

—— —— 'Errata the determinants of US direct investment in the EEC', *American Economic Review*, vol. 61, 1971

—— —— 'The determinants of US direct investment in the EEC Reply to comments by M. A Goldberg', *American Economic Review*, vol 62, 1972.

SCHMITZ, A, and BIERI, J 'EEC tariff and US direct investment', *European Economic Review*, vol. 3, 1972

SCHOLLHAMMER, H, *Locational Strategies of Multinational Corporations* (mimeo), 1972

SCHREIBER, J, *US Corporate Investment in Taiwan*, Harvard University Press 1970

SEVERN, A K., 'Investment and financial behaviour of American investors in manufacturing industry', in F Machlup, W Salant and L Tarshis (eds), *International Mobility and Movement of Capital*, National Bureau of Economic Research 1972.

SHARPE, W. F., 'Capital asset prices a theory of market equilibrium under conditions of risk', *Journal of Finance*, vol 19, 1964

SMITH, K V., and SCHREINER, J C., 'A portfolio analysis of conglomerate diversification', *Journal of Finance*, vol 24, 1969

SPITALLER. E, 'A survey of recent quantitative studies of long term capital movements', *I M F. Staff Papers*, Mar 1971

STEVENS, G V G, 'Fixed investment expenditure of foreign manufacturing affiliates of US firms theoretical models and empirical evidence, *Yale Economic Essays*, vol 9, Spring 1969

—— *Risk and Return on Selection of Foreign Investments* (mimeo), 1969

—— 'Capital mobility and the international firm', in F Machlup, W. Salant and L. Tarshis (eds), *International Mobility and Movement of Capital*, National Bureau of Economic Research, 1972.

—— 'The multinational enterprise and the determinants of investment' in J H Dunning (ed.), *Economic Analysis and the Multinational Enterprise* (forthcoming).

STOBAUGH, R. B., 'Where in the world should we put that plant?' *Harvard Busines Review*, Jan./Feb. 1968.

—— 'How to analyse foreign investment climates', *Harvard Business Reviеu* Sept./Oct 1969.

—— *The Multinational enterprise and the petrochemical industry*, Basic Books, New York, 1973.

STONEHILL, A. I., *Foreign Ownership in Norwegian Enterprise*, Oslo, Central Bureau of Statistics, 1965.

—— and NATHANSON, L., 'Capital budgeting and the multinational corporation' in A I Stonehill (ed.), *Readings in International Financial Management*, Good Yea Publishing Co., 1970.

STOPFORD, J. M., *Organising the Multinational Firm Can the Americans Learn fror the Europeans?* (mimeo), 1973

STRASSMAN, W. PAUL, *Technological Change and Economic Development*, Cornel University Press, 1968

STUBENITSKY, F., *American Direct Investment in the Netherlands Industry*, Rotterdan University Press, 1970

TILTON, J. E., *International Diffusion of Technology the Case of Semiconductors* The Brookings Institution, 1973.

TOBIN, J., 'Liquidity preferences as behaviour towards risk, *Review of Economi Studies*, vol 26, 1958

—— 'The theory of portfolio selection' in F. Hahn and F. P. R Brechling (eds.) *The Theory of Interest Rates*, Macmillan, 1965.

United Nations, *Economic Survey of Asia and the Far East*, Bangkok, 1971.

U S Department of Commerce, *US Direct Investments Abroad 1966*, Part II · Investment position, financial and operating data, Group 2 Manufacturing industries 1972

VAITSOS, C., *Intercountry Income Distribution and Transnational Corporation* (unpublished), 1972.

VAUPEL, J W., *Characteristics and Motivations of the US Corporations whic manufacture Abroad*, paper presented to meeting of participating members of the Atlantic Institute, Paris, June 1971. 1971.

VERNON, R., 'International investment and international trade in the product cycle' *Quarterly Journal of Economics*, vol 80, 1966.

—— *Sovereignty at Bay*, Basic Books, 1972

—— 'The multinational enterprise and the location of economic activity' in J. H Dunning (ed.), *Economic Analysis and the Multinational Enterprise*, Allen and Unwin (forthcoming).

WALLIS, K. F., 'Notes on Scaperlanda's article', *Economic Journal*, vol 73, 1968.

WEBER, A., 'Location theory and trade policy', *International Economic Papers*, 1958

WELLS, L. T., *The Product Life Cycle and International Trade*, Harvard University Press, 1972

—— *Economic Man and Engineering Man· Choice of Technology in a Low Wag Country* (mimeo), 1972.

WILKINS, M., *Emergence of Multinational Enterprise*, Harvard University Press, 1970

WILKINSON, B., *Canada's International Trade an Analysis of Recent Trends and Patterns*, Private Planning Association of Canada, Montreal, 1968

WOLFE, B., *Internationalisation of US Manufacturing Firms a Type of Diversification* (unpublished doctoral thesis, Yale), 1971.

WORTZEL, W. H., *The Multinational Enterprise and the Pharmaceutical Industry* Basic Books, New York, 1973.

HUMAN SKILLS, R AND D AND SCALE ECONOMIES IN THE EXPORTS OF THE UNITED KINGDOM AND THE UNITED STATES[1]

By HOMI KATRAK

I. Introduction

ALBERT HIRSCHMAN writes about the Principle of the Hiding Hand [9] to illustrate how the emergence of unforeseen difficulties in development projects stimulates a search for hitherto unthought-of solutions and new ideas Something like the Hiding Hand may also be observed in the field of international trade theories, for following the publication of W. W Leontief 's findings for the Heckscher–Ohlin theory [23] there has been a fruitful search for alternative models of trade (while some older theories have also been revived) Among the more interesting of these are the Human Skills, the Scale Economy and the Technological Gap theories [2] The Human Skills theory aims to explain trade patterns in terms of countries' endowments of human and physical capital on the one hand and unskilled labour on the other and industries' requirements of these factors. The Scale Economy theory points to the importance of national market size in influencing relative costs and exports Whereas both these theories assume that countries share in a given state of technological knowledge, the Technological Gap theory argues that such knowledge is neither given nor universally available but is acquired through the R and D efforts, this latter theory emphasizes the forces that influence countries to invent and innovate products and to export these during the time that they have a temporary monopoly in the production of such products.

Interest in these theories has been enhanced by an impressive output of empirical findings notable among which, and of most relevance to this paper, are (i) D. B. Keesing's [16] test of the Human Skills theory which suggests that trade patterns can be explained in terms of countries' labour-skill availabilities and industries' requirements of these factors,[3] (ii) G. C. Hufbauer's test of the Scale Economy theory [11] showing that exports of

[1] I am grateful to Professor H. G Johnson for comments on an earlier draft of the paper; several helpful comments are also owed to members of the Economics Seminar at the University of Surrey The research for the paper was made possible by a grant from the Faculty IV Research Fund, University of Surrey.

[2] The Human Skills theory originated with a suggestion of W W Leontief [23], and has since been rigorously formulated by P B Kenen [21] The Technological Gap theory is associated with M V Posner [26] and R Vernon [33]. The influence of scale economies on trade was suggested early on by B Ohlin [25] For a discussion of these theories see J. N Bhagwati [2], G C Hufbauer [11], and H G Johnson [12, 13] Interesting tests, other than those mentioned in the text are by J. N Bhagwati and R Bharadwaj [3] and P B Kenen [20]

[3] Keesing used the ratio of skilled/unskilled labour employed in an industry as an indicator of the industry's requirements of human and physical capital per unit of unskilled labour.

large countries are more intensive in scale elasticity parameters than are exports of smaller countries, and (iii) a test of the Technological Gap theory by W. Gruber, D Mehta, and R. Vernon [6] who found that the R and D efforts of United States industries explained that country's comparative exports *vis-à-vis* non-European countries, but not with other technologically advanced countries, e g. the United Kingdom and West Germany. Now although these researches have produced interesting results, some aspects of the theories as well as the procedures for testing them raise questions that are worth examining

As W. Gruber *et al* 's test of the Technological Gap theory took account of R and D in the United States only, it seems that they were concerned with product innovation in just one country, this procedure has a shortcoming pointed out by J N Bhagwati [2] the tests 'fall short of investigation of the R and D indices of partner countries'.[1] One purpose of this paper is to examine the implications for the theory and for its empirical testing of R and D activities in a pair of countries G C. Hufbauer's test of the Scale Economy theory also has a shortcoming in that the levels of output in individual industries in the countries concerned have been ignored, this paper undertakes an alternative test of the theory incorporating the industry production levels along with Hufbauer's scale elasticity parameters. Section II discusses some problems involved in testing the Technological Gap and Scale Economy theories, and Section III tests all the three theories with reference to U S /U K exports of major industries, the reason for choosing this pair of countries is partly due to availability of data, partly in order to examine the result of W. Gruber *et al* for this pair of countries, and also because the differences in factor costs and economic size of these countries make the pair suitable for a test of the Human Skills and Scale Economy theories. It is perhaps worth emphasizing that the purpose is to examine relative exports of various industries at a point of time, for although there has recently been much interesting work, particularly in relation to the Technological Gap Theory, analysing inter-temporal changes in the trade patterns of a particular industry,[2] and although answers to both questions are needed for a fuller understanding of international trade flows, each of these seems sufficiently important to merit separate study.

II. Some problems in the empirical testing

As a starting-point it is convenient to summarize the procedure of W. Gruber *et al.* for testing the Technological Gap theory, this may be described as follows.

$$E_i^a/E_i^b > E_j^a/E_j^b \quad \text{whenever } R_i^a > R_j^b, \tag{1}$$

[1] J. N. Bhagwati [2], p 109. [2] A. Harman [8] and G. C. Hufbauer [10]

where R denotes an industry's R and D intensity (measured alternatively as the R and D expenditure as a percentage of sales, and as scientists and engineers in R and D as percentage of total employment) E denotes exports to world markets,[1] superscripts a and b refer respectively to the United States and to the other country in the comparison, and subscripts i and j refer to the ith and jth industries Using a rank correlation test for nineteen industry groups their procedure did not satisfactorily explain U S /U K. exports. The question now is whether any better results can be obtained if the tests incorporate the R and D of the United Kingdom as well as that of the United States, some procedures for such a test are discussed below

Let us suppose that each country's R and D generates some products and processes that are different[2] from those resulting from the R and D in the corresponding industry in the other country For convenience, these products and processes that are available to only one country (by virtue of its R and D) will be called 'new products' Drawing on M V Posner's [26] and R. Vernon's [33] discussion of the Technological Gap theory we may expect that an industry's exports will be greater the greater its output of new products In turn this suggests that a country's exports will be comparatively greater in the industry in which its output of new products is comparatively greater.[3] The problem now is to translate this into a test in terms of the R and D intensities (measured as the R and D expenditures per unit of sales[4]) For this we need to know whether the impact of $1 R and D (per unit of sales) on new products and on exports is the same for all industries W Gruber *et al* seem to have assumed that it is—this is implicit in their procedure of ranking United States' R and D intensities against U S /U K exports Keeping to this assumption a test of the theory would be

$$E^a_i/E^b_i > E^a_j/E^b_j \quad \text{whenever } (R^a_i - R^b_i) > (R^a_j - R^b_j) \qquad (2)$$

where superscript b now refers to the United Kingdom An alternative procedure would be to assume that the effect of $1 R and D (per unit of sales) on new products and on exports may differ between industries but that the 'elasticity' of new products with respect to changes in the R and D

[1] They correlate 1962 exports and R and D of the same year and hence make no allowance for time lags between the spending of R and D funds and their impact on exports D B Keesing [16] did find that lagging exports two or three years behind R and D made little difference to the results

[2] Admittedly there is no precise meaning to saying that countries' exports are 'different', but the general idea is that (as far as the Technological Gap theory is concerned) we need to get away from the assumption that within a particular industry countries produce identical and homogeneous products

[3] This assumes that world demand is not biased as between the new products of the two countries

[4] Although Gruber *et al* [6] also used an alternative measure of R and D intensities, namely scientists and engineers in R and D as percentage of total employment, this will not be used here since inter-country differences in the employment mix may be influenced by relative differences in returns to different types of employees

intensity is the same for all industries in the two countries; the appropriate test would now be

$$E_i^a/E_i^b > E_j^a/E_j^b \quad \text{whenever } R_i^a/R_i^b > R_j^a/R_j^b. \tag{3}$$

Ideally the tests should incorporate time lags since R and D expenditures in a particular year may enable exports of new products in future years. But, unfortunately, practical considerations make this very difficult since the relevant lag may differ between industries. At the same time, there is no *a priori* reason to expect that neglect of time lags would cause a systematic bias in the testing procedure; it can be shown that a bias would arise only if the lag were always comparatively greater in the industry and country where the R and D intensity were comparatively greater.[1]

What role do comparative costs have in (2) and (3)? If output of new products is large in relation to that of the appropriate industry, the two countries' exports will have few close substitutes and comparative costs will have little influence on relative exports. Conversely, the smaller is new product output (in relation to total output of the industry) the greater will be the influence of comparative costs. And it may be possible that a country's exports will be comparatively smaller in the industry where its R and D is comparatively greater; in turn, this suggests the possibility that comparative costs may cause the Technological Gap theory to be empirically invalidated.

There is one other important problem which should be mentioned. Implicit in the discussion of the above procedure (as in the empirical analysis of Gruber *et al.*) is the assumption that products are desired for their own sake, i.e. goods, *per se*, give utility. However, an alternative view of consumer demand, suggested by K. J. Lancaster [**22**], is that consumers purchase products because of the characteristics contained in them; by buying goods in certain proportions consumers can obtain their desired bundle of characteristics at lowest cost. In his theory a new product is one that contains characteristics in different proportions from those contained in existing products. This line of reasoning allows an analysis [**15**] of product innovation and exports in terms of the characteristics of new and old exports; briefly the argument is as follows. Suppose that an industry in Country A exports a product that contains the same characteristics as that contained in some product exported by other countries, but suppose

[1] Suppose that of the total stream of exports that are generated by $1 of current year R and D, only Z per cent can be exported in the current year; Z is an indicator of the time lag, i.e. the higher is Z the shorter is the lag. Now if $Z_i^a = Z_i^b$ and $Z_j^a = Z_j^b$, the lags will have no effect on our testing procedure. Next suppose that $Z_i^a \neq Z_i^b$ and $Z_j^a \neq Z_j^b$; even in this case neglect of lags will not give misleading results unless $Z_i^a/Z_i^b < Z_j^a/Z_j^b$ whenever $R_i^a/R_i^b > R_j^a/R_j^b$, or alternatively whenever $(R_i^a - R_i^b) > (R_j^a - R_j^b)$. In other words the value of Z does not matter unless Z is always comparatively lower in the country and industry where R is higher.

also that A's export is relatively more intensive in characteristic Z_1, while the export of its foreign competitors is relatively more intensive in Z_2. Now suppose that the relevant industry in Country A innovates a product which, though containing only Z_1 and Z_2, is more intensive in Z_1 than A's initial export. In such case the new product export may completely replace the old product and, what is more important, the value of A's exports from that industry may fall below the level prevailing before the innovation, the reason for this result is that A's products are desired by foreign consumers because of the high Z_1 content and the innovation enables them to reduce their purchases from A and yet to satisfy their initial demand for Z_1. Of course, some other new products may cause an increase in exports—this may arise, for instance, if the new product is more intensive in Z_2 than the old product. But the over-all conclusion from [15] is that there is no *a priori* reason to expect that product innovation in an industry will increase its exports. The implication for the present tests is that there is no presumption that industries with high R and D will have relatively high new product exports.

In view of these theoretical considerations we should approach our empirical analysis of the Technological Gap theory in an agnostic manner. The results that we obtain can lead us to ask further questions—but at this stage we do not really have clear-cut expectations of the nature of these results.

Let us consider now the testing of the Scale Economy theory. The purpose here is to build upon the procedure of G. C. Hufbauer [11] which may be described as follows: if α_j is the scale elasticity of the jth industry[1], X_j is the proportion of a country's exports that are due to the jth industry, calculate $\Sigma_j \alpha_j X_j$ for several countries and see whether these are related to some index of country size, for a twenty-four nation test he obtained fairly satisfactory results using alternatively national manufacturing output, GDP, and GDP *per capita* as indicators of size.

But let us consider this procedure a little. The underlying reasoning is that large countries will have relatively large exports of industries subject to greater scale economies. We know that if the relative size of various industries were the same in any pair of countries and if each country's output of an industry were produced in a single plant, the larger country would have a comparative cost (and export) advantage in the industry with greater scale economies.[2] However, since the relative industry size and the

[1] The value of α is calculated by him from the equation $v = kn^{\alpha}$, where v is the 1963 ratio between value added per man in plants employing n persons and the average value added per man for the four digit U.S. Census Bureau industry, and k is a constant, if α = say 0.06, a doubling of plant size will increase output per man by roughly 6 per cent.

[2] This argument abstracts from all influences on comparative costs other than scale economies. Also I focus only on the effects of scale on costs of production and ignore the possibility that larger firms may undertake more R and D per unit output, some evidence justifying this assumption is to be found in J. W. Markham [22].

distribution of output within different sized plants may differ between countries, they should be taken into account in our test (along with the scale elasticities) Initially it was hoped to undertake a test on such lines; but, unfortunately, the frequency distribution of different sized plants in the United States and the United Kingdom are not easily comparable. It is, however, possible to take account of inter-country differences in the relative size of industries, and this we will do as follows

Let us suppose that each country's output of an industry is produced in a single plant (though the plant size may differ between countries and industries) Then for the U S /U K case we should expect that United States' cost advantage (and exports) will be relatively greater for the ith industry whenever

$$(N_i^a/N_i^b)^{\alpha_i} > (N_j^a/N_j^b)^{\alpha_j}, \tag{4}$$

where N_i^a is the level of employment[1] in the ith industry in the United States, and the other symbols have a corresponding meaning Hypothesis (4) is tested in the next section

In a recent paper R E. Baldwin [1] has also examined the effect of scale economies His interest was in explaining the structure of United States trade, and he used as a measure of scale economies in an industry the percentage of employees in United States establishments with 250 or more employees This procedure has an advantage over that suggested in (4) above in that it takes account of the relative importance of large and small plants in the United States On the other hand it ignores the estimates of scale elasticities, the industry output levels in the United States relative to those in other countries (and the distribution of different sized plants in the other countries), for this reason and because of the problem of poor comparability of plant size data mentioned above it was decided not to follow Baldwin's procedure (Of course economies of scale may be associated with large size of firm as well as large size of plant So ideally our tests should allow for the size-distribution of firms as well But unfortunately, as with the case of plants, such a test would be difficult because of problems of data comparability)

It is perhaps worth emphasizing that hypothesis (4) assumes that the 'industry' follows average, rather than marginal, cost pricing It is this assumption that makes it possible to infer the comparative cost of exports

[1] The argument here is that comparative costs are indicated by comparative output per man, and the latter are determined by the relative size of (the single plant) industries and the scale elasticities Industry size is being measured by employment rather than by output because Hufbauer's scale elasticities had been estimated with employment as an indicator of plant size There is a practical problem here in that whereas the hypothesis of trade patterns should be based on comparative costs and employment levels in the pre trade situation, the tests must of necessity be undertaken with data that pertain to the trading situation, as I have pointed out elsewhere [14], a similar problem arises in testing the Ricardian and Heckscher–Ohlin theories

from the data on comparative industry size The alternative assumption of marginal cost pricing would have to recognize that the level of exports of a particular industry may depend on the demand elasticities in the home and export markets, the lack of empirical data of such elasticities makes that procedure rather impracticable A further point is that because of economies of scale our procedure is logically permissible only if we assume some differentiation of products between the two countries' industries—in the absence of such conditions one country's industry would capture the entire world market

For the test of the Human Skills theory I will follow D B Keesing [16] in assuming that an industry's requirements of skilled/unskilled labour are a reliable proxy for its use of human and physical capital per unit of unskilled labour And then assuming that the United States is relatively well endowed with capital (in both its forms), the test will be that U S /U K. exports of individual industries are greater the greater is the skill intensity of the industry The only difference between this test and that of Keesing is that he was concerned with the skill intensities embodied in countries' aggregate exports and imports, while we are testing for a relationship between skill intensities and relative exports of individual industries [1]

There is, however, a practical problem here in that the dividing line between 'skilled' and 'unskilled' labour is likely to be arbitrary Keesing's earlier exercise [18] relied on a weighted index showing the employment of scientists, engineers, etc , relative to that of semi-skilled and unskilled employees It seems desirable to use both indices in the tests for this may enable us to see whether our results could be influenced by the choice of any particular index For the index used in [18] I used the data contained there to compile skill intensities in the United States, Keesing's formula is reproduced in the Appendix of this paper And from [19] we have data for skill intensities in the United States and in the United Kingdom

One qualification about the use of these data should be mentioned In so far as the Human Skills theory emphasizes the influence of relative factor endowments on comparative costs of production our data should pertain to the skilled/unskilled ratio of employees involved in production only The problem now is that our data cover all employees, i e. those in R and D as well as those in production—separate data for these categories are not available Fortunately this problem is unlikely to be very serious Data in Keesing [18] show that in several United States industries the numbers

[1] Alternatively, the theories could be tested against relative imports of individual industries However, in so far as the Technological Gap theory emphasizes the role of product innovation, this is likely to have a greater impact on a country's exports than on its import-competing output In addition, relative import patterns are likely to be influenced by the peculiarities of each country's tariff structure whereas their exports face rather similar barriers in foreign markets

of scientists and engineers in R and D are small in relation to the industries' employment of these categories; moreover he found that for eighteen industries the linear correlation between the number of scientists and engineers in R and D and those elsewhere in the industry was 0·66. Now since the major category of employees in R and D are likely to be scientists and engineers it seems reasonable to expect a significant correlation between the skilled/unskilled ratios being used in our tests and the ratio of skilled/unskilled employees in production only.

III. The empirical tests

A starting problem was to settle on the number of observations to be included. Since the industrial classification for the various data (R and D, skills, scale elasticities) are not always similar, some aggregation became necessary. If all the series are translated into a common classification we are restricted to only fourteen observations. However, we gain degrees of freedom if we work with two slightly dissimilar industrial classifications, one each for our two measures of skill intensities we have seventeen observations for the United States' skill data compiled from [18] and fourteen observations when working with the skill data (of both countries) from [19], the data for R and D, scale elasticities, and exports can be telescoped to fit the seventeen and fourteen observations Details of the industrial classification and sources are in the Appendix

The initial test was to check for the possibility of skill intensity reversals between the two countries, this can be done for the fourteen observations obtained from [19] The rank correlation between the two countries' skill intensities is 0·97 (significant at 1 per cent)

It would be desirable to see whether similar results also arise with the other measure of skill intensity, however, due to unavailability of United Kingdom data for that measure the relevant test cannot be undertaken All we can say is that for the only test permitted by data availability we find little evidence of skill intensity reversals as between the two countries A further question is whether the ranking of industries (by skill intensity) within any one country depends on the choice of the index of skill intensity To examine this I telescoped the seventeen observations from [18] to fit the fourteen observations from [19], thus obtaining two measures of skill intensities in identical industries in the United States; the rank correlation between these is 0·91 (significant at 1 per cent) This suggests that the performance of the Human Skills theory is unlikely to differ as between the two measures of skill intensity

The next step was to examine for a possible interrelationship between the forces emphasized by the three theories. Each of the theories offers an explanation of U.S./U.K. exports of individual industries; the human

Skills and Scale Economy explanations run in terms of the influence on comparative costs (though the explanation differs as between the two theories), while the Technological Gap explanation focuses on relative output of new products We wish to see whether the theories suggest similar patterns of United States' export advantage. To examine this six pair-wise rank correlations were calculated (for each of the two industrial classifications) between the two measures of the R and D gap, the skill intensities[1] and our index for the effect of scale economies, i e $(N_j^a/N_j^b)^{\alpha_j}$

TABLE I

Rank correlations between industry characteristics

	$(R_j^a - R_j^b)$	R_j^a/R_j^b	*Skill intens*	*Scale effects*
Seventeen industries classification a				
$(R_j^a - R_j^b)$	1 00	0 74*	0 52*	0 13
R_j^a/R_j^b		1 00	0 29	−0 37
Skill intens.			1 00	0 15
Scale effects				1 00
Fourteen industries classification b				
$(R_j^a - R_j^b)$	1 00	0 80*	0 67*	0 21
R_j^a/R_j^b		1 00	0 39	−0 26
Skill intens.			1 00	0 36
Scale effects				1 00

See notes to Table II.

The results are shown in Table I All the correlations except that between the two measures of the R and D gap are fairly poor, this is interesting for two reasons First, the empirical performance of the Technological Gap theory is unlikely to differ as between our two measures of the R and D gap, i e as between our alternative assumptions about the relationship of R and D and new product output. Secondly, the weak rankings between the forces of the three theories suggest that these forces may occassionally conflict with each other and give contrary predictions of relative export performance One aspect of the latter conclusion is perhaps worth emphasizing Within a particular country the industries that have high R and D expenditures may also have relatively high proportions of scientists, engineers, etc , in production However, these similarities in industry characteristics do not make the Technological Gap and Human Skills theories merely different versions of a common explanation of trade; nor need they yield similar predictions of trade patterns. The point is that

[1] United States' skill intensities are used in tests with the seventeen industries classification and United Kingdom intensities are used for the fourteen industries classification

whereas the predictions of the Human Skills theory, because of its explicit assumption of international identity of production functions, can be derived with reference to the skill intensity data of either country,[1] *the Technological Gap theory must take account of the R and D in both the countries involved*

Another initial problem was to decide on which year's export data to use in the tests. W Gruber *et al* [6] ranked United States' R and D in 1962 against U.S /U.K exports in the same year. However, it may be argued that a later year's export data should be used because there may be a time

TABLE II

Rank correlations between 1962 U S /U K exports and industry characteristics

Number of industries	$(R_j^a - R_j^b)$	R_j^a/R_j^b	*Skill intens*	*Scale effects*
17 *a*	0 42*	−0 11	0 56*	0 59*
14 *b*	0 22*	−0 16	0 37*	0 76*

Notes R_j^a and R_j^b are the R and D intensities in the United States and the United Kingdom respectively. The Scale effect is the coefficient $(N_j^a/N_j^b)^{\alpha_j}$. The sources of data and other details are in the Appendix.

a Using skill intensities of the United States.

b Using skill intensities of the United Kingdom

* Indicates significant at 5 per cent

lag between the R and D expenditures and their effect on exports Another reason for using some later year's export data is that the data on R and D, skills, etc , stretch over the period 1962–5 In order to see whether choice of year for the export data could have an important bearing on the results I ranked U S./U K exports in 1962 with those in 1964 and in 1966. For the seventeen industry classification the correlations are 0·97 and 0·94 for 1962/4 and 1962/6 respectively, the corresponding results for the fourteen industry classification are 0·96 and 0·91 respectively. (All results are significant at 1 per cent) This suggests that our tests may equally well make use of 1962, 1964, or 1966 export data It was decided to keep with W Gruber *et al* and use 1962 data for all of the tests; and in addition we can check some of the results with 1964 and 1966 data.

The three theories were tested first by rank correlations and then by multiple regressions In each case we use the seventeen and fourteen industries classifications. For the former (latter) classification we use the skill intensities of the United States (United Kingdom). The rank correlations were undertaken with 1962 export data only; the results are shown in Table II The over-all impressions are as follows. The R and D explanation does

[1] Except when there are skill intensity reversals in which case separate (and possibly conflicting) predictions will be derived for the intensities of each country.

not perform well in either of the alternative measures The Human Skills theory has a partial success the correlation with the more detailed classification is significant at 5 per cent The Scale Economy explanation comes off best of all (under both the industrial classifications) We may recall that the scale coefficient $(N_j^a/N_j^b)^{\alpha_j}$ purports to show the influence of scale elasticities and of comparative industry size on comparative costs and hence on relative exports; for convenience this coefficient will be called the scale effects

The multiple regressions were undertaken in the linear and log-linear form 1962 export data were used for both the industrial classifications and in addition 1964 and 1966 data were also used for the more detailed classification The results are in Table III It turned out that the results differ little as between the two measures of the R and D gap and for this reason the table reports only those for the measure $(R_j^a - R_j^b)$ Considered together the theories perform rather well the R^2 values (corrected for degrees of freedom) range from 0·61 to 0 91, the F-statistic in all cases is significant at 5 per cent The results for 1962 and 1964 are very similar and somewhat more favourable than those for 1966 Turning to the results for each of the theories the story is much the same as with the rank correlations The Technological Gap theory comes off poorly in all the regressions, the t-values for its coefficient are not significant at 5 per cent This result (and the rank correlation) are in line with that of W Gruber *et al* , what the present tests add is that incorporation of the R and D of the United Kingdom, as well as that of the United States, does not improve the performance The results for the Human Skills theory are more satisfactory (though its coefficient does not reach 5 per cent significance in all the cases) Here we are in line with D B Keesing's [**16**] results for the Human Skills theory, in a sense our results for the industry-by-industry test are more reassuring than his findings since he was concerned only with the skill requirements of aggregate exports and imports The most encouraging of our results are for the Scale Economy theory, in all the regressions the coefficient is significant at 5 per cent and in some cases even at 1 per cent [1] Since our procedure of using the scale effect coefficient has no precedent, comparisons with earlier work are not possible

IV. Some further questions

A number of questions may arise about the interpretation of the above results First let us consider possible questions about the Scale Economy tests It may be argued that since our tests for that theory incorporate relative industry size the results could merely be showing that industries

[1] As with the rank correlations the multiple regression shows little inter relationship between the forces of the three theories

with relatively large size are also those with relatively large exports; in other words the results may indicate little about the influence of scale economies on exports. This problem may be examined by testing whether

TABLE III

Regression equations relating U S /U K exports to comparative differences in R and D, skill intensities, and scale effects

Seventeen industries classification *a*

Linear equations

$$X_{62} = -1711{\cdot}94 + \underset{(0\ 26)}{0{\cdot}17R} + \underset{(0\ 74)^*}{1\ 36H} + \underset{(2\ 98)^*}{17\ 23N}; \quad R^2 = 0{\cdot}896$$

$$X_{64} = -1275\ 00 + \underset{(0{\cdot}19)}{0\ 07R} + \underset{(0\ 54)^*}{1\ 30H} + \underset{(2\ 19)^*}{13\ 20N}, \quad R^2 = 0\ 906$$

$$X_{66} = -165{\cdot}45 + \underset{(0\ 13)}{0{\cdot}09R} + \underset{(0\ 38)}{0\ 47H} + \underset{(1{\cdot}51)^*}{2\ 73N}, \quad R^2 = 0\ 645$$

Log-linear equations

$$X_{62} = -9\ 73 + \underset{(0\ 06)}{0{\cdot}02R} + \underset{(0\ 07)^*}{0\ 17H} + \underset{(1{\cdot}00)^*}{5\ 75N}; \quad R^2 = 0\ 842$$

$$X_{64} = -7\ 71 + \underset{(0\ 05)}{0\ 004R} + \underset{(0\ 06)^*}{0{\cdot}19H} + \underset{(0\ 89)^*}{4\ 76N}; \quad R^2 = 0\ 840$$

$$X_{66} = -2\ 27 + \underset{(0\ 06)}{0\ 04R} + \underset{(0\ 07)^*}{0\ 16H} + \underset{(1\ 02)^*}{2\ 03N}; \quad R^2 = 0\ 609$$

Fourteen industries classification *b*

Linear equations

$$X_{62} = -6200{\cdot}32 - \underset{(0\ 81)}{0\ 07R} + \underset{(2{\cdot}26)}{3{\cdot}97H} + \underset{(17\ 99)^*}{71\ 39N}, \quad R^2 = 0\ 700$$

Log-linear equation

$$X_{62} = -9\ 15 - \underset{(0\ 04)}{0\ 02R} + \underset{(0\ 06)^*}{0\ 12H} + \underset{(1\ 47)^*}{6{\cdot}02N}, \quad R^2 = 0\ 707$$

Notes: R denotes the measure $(R_j^a - R_j^b)$, i.e. the R and D gap between the United States and the United Kingdom, H denotes the skill intensity, N denotes the scale effect $(N_j^a/N_j^b)^{\alpha_j}$, X_{62} denotes U S./U K. exports in 1962, etc.

The standard errors of the coefficients are shown in parentheses

* Indicates t-values significant at 5 per cent

a United States skill intensities were used in these regressions

b United Kingdom skill intensities were used in these regressions.

relative exports can be explained in terms of industry size only, i e without reference to the scale elasticities. Table IV shows the results of such tests in the multiple regressions we now use the relative industry size (N_j^a/N_j^b) rather than the scale effect $(N_j^a/N_j^b)^{\alpha_j}$; the other explanatory variables are (as previously) the skill intensities and $(R_j^a - R_j^b)$. Comparisons with the results in Table III show that, for both the seventeen industry and fourteen

industry observations,[1] the relative industry size performs considerably less satisfactorily than the scale effects.

A related question is whether the procedure of using the scale effects does better than merely using the scale elasticities. In order to examine this

TABLE IV

Regression equations relating U.S./U.K. exports to comparative differences in R and D, skill intensities, and industry size

Seventeen industries classification *a*

Linear equations

$$X_{62} = -1258.22+0.78R+1.93H+0.39E, \quad R^2 = 0.654$$
$$\qquad (0.45) \quad (1.34) \quad (0.38)$$

$$X_{66} = 45.66+0.23R+0.42H+0.23E, \quad R^2 = 0.676$$
$$\qquad (0.11)^* \quad (0.37) \quad (0.10)^*$$

Log-linear equations

$$X_{62} = -0.54+0.15R+0.30H+0.36E, \quad R^2 = 0.526$$
$$\qquad (0.09) \quad (0.11)^* \quad (0.24)$$

$$X_{66} = 1.18+0.10R+0.19H+0.22E, \quad R^2 = 0.518$$
$$\qquad (0.06) \quad (0.07)^* \quad (0.15)$$

Fourteen industries classification *b*

Linear equations

$$X_{62} = 56.61+0.16R+0.23H+0.20E, \quad R^2 = 0.444$$
$$\qquad (0.11) \quad (0.35) \quad (0.10)$$

Log-linear equations

$$X_{62} = 1.19+0.07R+0.13H+0.26E, \quad R^2 = 0.287$$
$$\qquad (0.07) \quad (0.10) \quad (0.18)$$

Notes. R denotes the measure $(R_j^a - R_j^b)$, i.e. the R and D gap between the United States and the United Kingdom, H denotes the skill intensity, E denotes the relative output level (N_j^a/N_j^b), X_{62} denotes U.S./U.K. exports in 1962, etc.

The standard errors of the coefficients are shown in parentheses.

* Indicates t-values significant at 5 per cent.

a United States skill intensities were used in these regressions.

b United Kingdom skill intensities were used in these regressions.

the initial tests were repeated but now using α_j instead of the scale effects. The results given in Table V show that the R^2 values are considerably lower than with the scale effects and in three of the six regressions the t-value for the elasticity coefficient is not significant at 5 per cent. Considered all together these results show that whereas the combined influence of industry size and scale elasticities—as captured in the scale effects—does provide a significant explanation of relative exports, neither industry size *per se* nor scale elasticities *per se* seem to have much influence.

[1] 1964 was excluded from this test for as seen earlier the results for that year are likely to be very similar to that for 1962.

The next question concerns the testing of the Human Skills theory In adopting Keesing's procedure we are assuming that industries with relatively high labour-skill intensities are those with high requirements of human and physical capital per unskilled labour and then examining

TABLE V

Regression equations relating U S /U K exports to comparative differences in R and D, skill intensities, and scale elasticities

Seventeen industries classification a

Linear equations

$X_{62} = -65\ 51 - 0\ 39R + 2\ 20H + 1\ 75E,\quad R^2 = 0\ 847$
$\qquad (0\ 45)^*\ (0\ 70)^*\ (0\ 37)^*$

$X_{66} = 90\ 03 + 0\ 13R + 0\ 56H + 0\ 17E,\quad R^2 = 0\ 582$
$\qquad (0\ 15)\ (0\ 40)\ (0\ 19)$

Log-linear equations

$X_{62} = 1\ 20 + 0\ 11R + 0\ 30H + 0\ 15E,\quad R^2 = 0\ 537$
$\qquad (0\ 09)\ (0\ 11)^*\ (0\ 09)$

$X_{66} = 1\ 54 + 0\ 07R + 0\ 21H + 0\ 09E,\quad R^2 = 0\ 566$
$\qquad (0\ 06)\ (0\ 07)^*\ (0\ 06)$

Fourteen industries classification b

Linear equation

$X_{62} = 584\ 59 - 0\ 99R + 3\ 60H + 7\ 44E,\quad R^2 = 0\ 588$
$\qquad (1\ 14)\ (2\ 65)\ (2\ 52)^*$

Log-linear equation

$X_{62} = 3\ 19 + 0\ 02R + 0\ 05H + 0\ 16E,\quad R^2 = 0\ 474$
$\qquad (0\ 06)\ (0\ 09)\ (0\ 07)^*$

Notes R denotes the measure $(R_j^a - R_j^b)$, i e the R and D gap between the United States and the United Kingdom, H denotes the skill intensity, E denotes the scale elasticity α_j, X_{62} denotes U S /U K exports in 1962, etc

The standard errors of the coefficients are shown in parentheses.

* Indicates t-values significant at 5 per cent

a United States skill intensities were used in these regressions

b United Kingdom skill intensities were used in these regressions

whether United States' comparative advantage is in industries intensive in capital (in both its forms) However, a problem of interpretation arises here For example suppose that (i) an industry's employment of skilled labour were strictly complementary to its requirements of physical capital and (ii) that long-run average costs of the former were small in relation to those of physical capital and of unskilled labour In such case, the ranking of industries by their skill intensities may approximate that by the physical-capital/labour ratios, and our test of the Human Skills theory may in fact be a test of the conventional Heckscher–Ohlin theory (which emphasizes industries' requirements of physical capital and unskilled labour) To

consider this problem I calculated physical-capital/labour ratios for the seventeen industries (of our earlier tests) from detailed data[1] given in G C Hufbauer [11] Ranking these ratios against the skill intensities gives a correlation of 0·08 which suggests that the Human Skills and Heckscher–Ohlin theories are unlikely to give similar predictions of relative exports or to show similar results in the empirical tests The physical-capital/labour ratios were then used in multiple regression analysis in place of the skill intensities.[2] Linear and log-linear regressions were tried with 1962 and 1966 export data but in all four cases the t-values of the physical-capital/labour coefficient was not significant at 5 per cent In one sense this result is nothing new for it is in line with earlier refutation of the Heckscher–Ohlin theory by W W Leontief [23] and other researchers Comparison with the result for skill intensities (in Table III) does suggest, however, that an analysis of international trade flows in terms of industry's requirements of skilled/unskilled labour and countries' relative endowments of labour skills may be more meaningful than the earlier analysis in terms of physical capital and labour alone

One other aspect worth examining is about a possible interrelationship between the forces of the Human Skills and Scale Economy theories Although we have seen that the correlations between the skill intensities and the scale effects are quite weak, it is still worth checking for some other ways in which those forces may be related A possible check is to see whether an industry's employment mix of various labour skills varies with its size, if the employment of scientists, engineers, etc , increases with size of plant, it may be that the United States' advantage in the skill intensive industries is partly due to the larger scale of production in that country [3] Unfortunately, the relevant data for such an exercise are not available Something, however, may be learnt by using a poor substitute we can compile the ratio of skilled/(semi-skilled+others) employed in 'small' and 'large' plants in each of eleven United Kingdom industries and then examine whether these ratios are greater in the 'large' plants [4] Of course these ratios, which are shown in Appendix Table A V, say nothing about the relative importance of scientists, engineers, and technicians[5] in different sized plants

[1] The data is described in [11] as capital per man "measured in U S dollars of approximately 1963 vintage, refers to fixed plant and equipment immediately employed in making the commodity, as produced in the United States" (p 221)

[2] The other two independent variables in the regressions were the series $(R^a_j - R^b_j)$ and $(N^a_j/N^b_j)^{\alpha_j}$.

[3] It seems appropriate to consider this relationship with reference to size of plant rather than size of industry, for if an increase in size of industry were to involve merely a duplication of existing plants there would be no change in the labour mix

[4] The dividing line between 'large' and 'small' plants is an employment level of 500 and over

[5] The source of these data [29] groups together the technical, administrative, and clerical categories, and for this reason the technical category had to be excluded from the test

and for this reason are not comparable with the skill intensity data used our previous tests (*nor for that matter are the data appropriate for* testi the Human *Skills theory*[1]) *None the less these ratios may give* som indication *of how the labour mix varies with plant* size. *We find* littl indication that the labour mix differs as between the two plant sizes th rank correlation between these ratios for 'small' and 'large' plants is 0 9; showing that industries with relatively skill-intensive 'small' plants ar also those with skill-intensive 'large' plants; and the mean of these intensi ties turned out to be slightly greater for the 'small' plants (the t-test for nul difference between the means was rejected at 5 per cent) In view of th nature of these ratios and the limited number of observations too mucl reliance cannot be placed on the results, none the less they do suggest tha the skills used do not increase with plant size and this supports our earlie finding that the forces of the Human Skills and Scale Economy theorie are independent of each other

A final question concerns the poor results for the Technological Ga theory. When W. Gruber *et al* [6] found that the R and D of the Unite States' industries did not explain U S /U K exports they suggested tha this was because both countries derive their competitive strength from th same set of forces, viz R and D and product innovation [2] In contrast, th argument of this paper has been that even though both countries are activ in R and D, it may still be possible to explain relative exports in terms o the differences in their R and D provided that the two countries are not innovating similar products; but yet our tests incorporating measures o the R and D gap have proved unsatisfactory

At first encounter it may seem that a likely explanation of the results i the neglect of time lags (between R and D expenditures and the effect on exports) But let us examine this a little Consideration of the time lags may begin by noting that in most industries R and D is not a once-and-for all expenditure but an annually recurring one Now if these recurring expenditures were such that the inter-country R and D gap (for each industry) remained stable over time, and if the effect of $1 R and D on new product output were also fairly constant, the two countries' relative exports of new products would be fairly stable from year to year, the question of time lags could be ignored—in fact, the problem would not exist—and we could correlate the R and D data of any year with the export data of that, or some later, year. The empirical question now is about how

[1] Out of curiosity I used these ratios in a multiple regression along with the $(R_j^a - R_j^b)$ and $(N_j^a/N_j^b)^{\alpha_j}$, the regression coefficient for the R and D and skills ratio are not significant at 5 per cent

[2] They write 'the United Kingdom and Germany, also being at the top of the advanced country list with relatively high incomes and a relatively heavy stress on industrial innovation and product development, derive their export strength from roughly the same characteristics as those that govern U S export performance' (p. 26)

the R and D gap behaves over a period of years Unfortunately since annual data for United Kingdom industries' sales are not available we cannot calculate the relevant R and D intensities, i e R_j^b, etc However, something may be learnt from the data on the absolute level of R and D expenditures The earliest data for the United Kingdom relate to 1961/2, these have a more aggregate classification and narrower definition than the 1964/5 data used in the tests above,[1] but fortunately we can obtain data for 1964/5 and for 1966/7 on the same basis as for 1961/2. United States' R and D expenditures for the same classification were collected for 1960, 1962, 1964, and 1966 The sources are [30] and [32] Using these data I compiled two sets of annual series of the R and D gap, the first subtracts (for each industry) United Kingdom expenditures in 1961/2 from United States expenditure in 1960, and correspondingly 1964/5 from 1962 and 1966/7 from 1964, the other set subtracts United Kingdom expenditure in 1961/2 from United States expenditure in 1962 and correspondingly for the later years By having two sets we add to the sensitivity of our tests This enables three pair-wise correlations for each of the two sets. The results of the six rank correlations range from 0·97 to 0 99; thus if the United States has a large (small) lead in R and D in a particular industry in any one year it is likely to have a corresponding large (small) lead in later years The implication of these results is fairly clear As long as R and D expenditures are appropriate indicators of new product output, we may infer that the ranking of industries by U S /U K output and exports of new products is likely to have remained fairly stable in the early 1960s, new product exports may lag behind R and D expenditures, but none the less any one year's R and D would explain the relative exports of new products of that year and of later years (Perhaps an alternative method of examining the effect of time lags would have been to see whether past growth rates of R and D had been the same in the two countries, but unfortunately as there is little information on R and D for individual industries for years prior to 1962 such a test is not possible)

One other explanation of our results should be mentioned Our tests of the Technological Gap theory have been solely in terms of R and D expenditures, the aim has been to build upon the work of W Gruber *et al* and to see if something could be learnt by incorporating the R and D of both countries However, it may be that R and D data are not an appropriate indicator of new product output even for a single industry,[2] for what matters is not

[1] The R and D data for 1961/2 cover expenditure within private industry only and exclude expenditure outside the unit These data permit thirteen observations comparable to United States R and D data

[2] An alternative indicator may be patent statistics, however, even these may be inappropriate for, as pointed out in [4], patent statistics may be more relevant as indicators of innovative input than of the output.

merely the production of new technology but also the use of it. It may also be that the level of aggregation involved in the tests—made inevitable by data availability considerations—was too broad to pick up the Technological Gap influences on relative exports; thus R and D may have a significant impact on exports but the effects may be restricted to only a few of the commodities in each of the industry headings

Further empirical research with more detailed data may perhaps yield more significant results for the Technological Gap theory For the present we may also recognize that the results could reflect the argument in Section II above that even at the theoretical level the relationship between product innovation and exports may be uncertain· if new products are desired not for their own sake but for the characteristics they contain it may be possible that product innovation may actually decrease exports of the innovating industry

It is worth emphasizing that although the aim of this paper has been to examine relative exports of a pair of countries, and although the procedure here has shown unfavourable results for the Technological Gap theory, that theory may still give considerable insight into other aspects of international trade Building on the analysis of M. V. Posner [**26**], G C Hufbauer [**10**] has shown how technical innovation in synthetic materials has shifted the location of production over time. innovation and initial production usually take place in advanced countries and they export proportionately more of the products of recent technology, while older products feature comparatively more in the exports of low-wage countries More recently, G F Ray [**27**] has investigated the diffusion of new industrial processes in a number of countries the Technological Gap theory may be of interest here also for it may be that inter-country differences in rates of diffusion of new technology are reflected in their export patterns

V. Summary

This paper has been concerned with an analysis of U S /U K. exports in terms of the forces of the Human Skills, the Technological Gap and the Scale Economy theories For the Human Skills theory I followed D B Keesing's [**16**] procedure of using industries' skill intensities as indicators of human and physical capital input per unit of unskilled labour for the Technological Gap theory I extended the procedure of W Gruber *et al* [**6**] using the R and D of the United Kingdom as well as of the United States and for the Scale Economy theory I built upon the procedure of G C Hufbauer [**11**], making use of industry output levels in the two countries as well as his estimates of scale elasticities Rank correlations and multiple regression analysis gave very satisfactory results for the Scale Economy theory

it was also found that the procedure of incorporating the relative output levels in the two countries as well as the scale elasticities gave better results than tests with the latter alone The results for the Human Skills theory were quite favourable, it was also shown that industries' skill intensities are not correlated with the physical-capital/labour ratios and that the latter do not provide a satisfactory explanation of relative exports The Technological Gap theory showed a disappointing performance, the effect on these results of time lags between R and D expenditures and the impact on exports was also considered but it was suggested that this may not have been a serious problem since the gap between the two countries R and D expenditures remained rather stable over the years involved in the tests

University of Surrey

APPENDIX NOTES

THE United Kingdom's R and D classification and data are from Statistics of Science and Technology [**30**], the Census of Production categories and output data are from Report on the Census of Production, 1963 [**28**] United States' SIC categories and R and D data were taken from W. Gruber *et al* [**6**], the correspondence between these categories and the United Nations' SITC is from [**6**] also The United Kingdom's R and D data are for 1964/5, while the census data are for 1963, the United States' data are for 1962

Skill intensities for United States were calculated from D B Keesing [**18**], this gives data of employment by occupational categories (in 1960) for a large number of SITC headings which correspond very closely (though not entirely) to those shown in Appendix Table A I Following Keesing the formula used to calculate skill intensities is

$$\frac{2(\mathrm{I}+\mathrm{II}+\mathrm{III})+\mathrm{V}}{\mathrm{VIII}}$$

where the skill classes are I = scientists and engineers, II = technicians and draughtsmen, III = other professionals, V = machinists, electricians, and tool-and-diemakers, VIII = semi-skilled and unskilled workers

The formula for the scale effects is $(N_j^a/N_j^b)^{\alpha_j}$, where N_j^a and N_j^b are the employments in the jth industry in the United States and United Kingdom respectively, and α_j is the scale elasticity. United States and United Kingdom data were taken from United States Census of Manufactures, 1963 [**31**] and Report on the Census of Production, 1963 [**28**] respectively. The values of α_j were obtained from G C Hufbauer [**11**], in some cases these are available for a more detailed classification than used in this paper and had to be averaged to our more aggregate classification (by using the weights used by Hufbauer, viz. United States' 1965 exports).

The data for the calculation of United Kingdom skill intensities is from Statistics on Incomes, Prices, Employment and Production, March 1965 [**29**] The dividing line between large and small plants is an employment level of 500 and over. The formula used to calculate skill intensities was

$$\frac{\text{skilled labour}}{\text{mainly semi-skilled and others}}.$$

APPENDIX TABLE A I

Assumed correspondence between United Kingdom and United States industrial classifications and United Nations S I.T C.

U K R D classification	*U.K census of production classification*	*U S SIC*	*U N S I.T C.*
Aircraft	64	372	734
Motor vehicles and railway equipment	62, 63, 65, 66	37 excl 372	73 excl 734
Electrical engineering	55 to 60	36	72
Scientific instruments	53	38	86
Chemicals and allied products	25 to 31, 33 to 36	28	5
Mechanical engineering	41, 42, 44 to 52	35	71
Rubber and rubber products	119	30	62, 893
Glass, pottery, building materials	102 to 107	32	661 to 666
Petroleum products	22 to 24	29	332
Metal products	43, 56, 68 to 72, 74	34	69
Non-ferrous metals	40	333	68
Iron and steel	37 to 39	33 excl 333	67
Clothing, footwear, leather	90 to 101	23, 31	83 to 85, 611 to 613
Paper, printing, publishing	114 to 118	26, 27	64, 892
Food, drink, tobacco	7 to 21	20, 21	013, 023, 024, 032 046, 047, 048, 053 055, 061, 062 091 099, 111, 112, 122
Textiles	75 to 89	22	65
Timber and furniture	109, 111 to 113	24 (excl 241, 242) 25	63, 243, 82

For notes and sources of data see Appendix notes.

Appendix Table A II

R and D, skill intensities, scale effects, and scale elasticities

Industrial classification	*R and D U.K.*	*U.S.*	*Skill intensit*	*Scale effects*	*Scale elast.*
Aircraft	23.02	27.2	1.69	1.35	0.304
Motor vehicles and railway equipment	1.52	2.8	0.40	1.04	0.054
Electrical engineering	5.70	7.3	0.73	1.05	0.063
Scientific instruments	3.97	7.1	1.04	1.04	0.038
Chemical and allied products	2.68	3.9	0.88	1.04	0.051
Mechanical engineering	1.69	3.2	0.80	1.02	0.044
Rubber and plastic products	1.34	1.4	0.28	1.03	0.030
Glass, pottery, and building materials	1.13	1.1	0.20	1.04	0.048
Petroleum products	1.74	0.9	1.00	0.98	−0.014
Metal products	0.46	0.8	0.57	1.02	0.028
Non-ferrous metals	0.64	0.8	0.39	0.94	−0.079
Iron and steel	0.68	0.5	0.29	1.05	0.069
Clothing, footwear, and leather	0.14	0.2	0.03	0.93	−0.058
Paper, printing, publishing	0.20	0.1	0.49	1.10	0.068
Food, drink, tobacco	0.29	0.2	0.10	1.09	0.093
Textiles	0.57	0.2	0.07	1.00	−0.001
Timber and furniture	0.38	0.1	0.07	1.07	0.052

For description of data see Appendix notes.

Appendix Table A III

Scientists, engineers, and technicians (except health) per 1,000 persons employed in the United States and United Kingdom

Industrial classification	*U.S*	*U.K*
Motor vehicles, railway equipment, and aircraft	94	86
Electrical engineering	121	108
Chemical and allied products	121	98
Mechanical engineering	73	97
Rubber and rubber products	41	36
Glass, pottery, and building materials	31	26
Petroleum products	98	84
Metal products	74	62
Primary metals	40	46
Clothing, footwear, and leather	6	10
Paper, printing, publishing	17	13
Food, drink, tobacco	13	17
Textiles	11	9
Timber and furniture	8	6

Source: D. B. Keesing [19].

Appendix Table A IV

U S/U.K. comparative exports to world markets

Industrial classification	*U.S./U.K. Exports*		
	1962	*1964*	*1966*
Aircraft	10·53	8·47	2·93
Motor vehicles and railway equipment	1·05	1·14	1·41
Electrical engineering	1·63	1·83	1·90
Scientific instruments	2·30	2·14	2·30
Chemical and allied products	1·92	1·98	2·00
Mechanical engineering	1·67	1·91	1·85
Rubber and rubber products	1·20	1·43	1·53
Glass, pottery, and building materials	1·14	1·15	1·21
Petroleum products	1·30	1·50	1·34
Metal products	1·08	1·15	1·40
Non-ferrous metals	0·98	1·20	1·01
Iron and steel	0·84	1·10	0·92
Clothing, footwear, and leather	0·83	0·90	1·00
Paper, printing, publishing	2·02	2·19	2·37
Food, drink, tobacco	1·39	1·27	1·05
Textiles	0·71	0·73	0·73
Timber and furniture	1·83	1·89	2·20

Sources. O E.C D Foreign Trade Statistical Bulletin, Series B, 1962, 1964, and 1966.

Appendix Table A V

Skill intensities in United Kingdom industries

Industrial classification	*Large plants*	*Small plants*	*All plants*
Vehicles (a)	0·69	1·28	0·76
Engineering and electrical goods (b)	0·54	1·06	0·71
Chemicals and allied industries (c)	0·43	0·34	0·42
Rubber and plastic products	0·34	0·35	0·34
Glass, pottery, building materials	0·51	0·57	0·56
Manufacture of metal goods	0·65	0·72	0·43
Clothing, footwear, leather	1·62	2·55	2·50
Paper, printing, publishing	1·18	3·36	2·21
Food, drink, tobacco	0·22	0·31	0·27
Textiles	0·38	0·46	0·56
Timber and furniture	1·01	1·56	1·49

(a) Including aircraft
(b) Including mechanical engineering and scientific instruments.
(c) Including petroleum products.
For notes and sources of data see Appendix notes

REFERENCES

1. BALDWIN, R. E., 'Determinants of the commodity structure of U.S. trade', *American Economic Review*, 1970.
2. BHAGWATI, J. N., 'The pure theory of international trade: a survey', in *Trade, Tariffs and Growth*, 1969.
3. —— and BHARADWAJ, R., 'Human capital and the pattern of foreign trade: the Indian case', *Indian Economic Review*, 1967.
4. COMANOR, W. S., and SCHERER, F. M., 'Patent statistics as a measure of technical change', *Journal of Political Economy*, 1969
5. FINDLAY, R., *Trade and Specialisation*, 1970.
6. GRUBER, W., MEHTA, D., and VERNON, R., 'The R and D factor in international trade and investment of United States industries', *Journal of Political Economy*, 1967.
7. —— and VERNON, R., 'The technology factor in a world trade matrix', in R. VERNON, ed. [34], below
8. HARMAN, A., *The International Computer Industry*, 1971.
9. HIRSCHMAN, A. O., *Development Projects Observed*, 1967.
10. HUFBAUER, G. C., *Synthetic Materials and the Theory of International Trade*, 1965.
11. —— 'The impact of national characteristics and technology on the commodity composition of trade in manufactured goods', in R. Vernon, ed. [34], below.
12. JOHNSON, H. G., *Comparative Cost and Commercial Policy Theory for a Developing World Economy*, Stockholm, 1968.
13. —— 'The state of theory in relation to the empirical analysis', in R. Vernon, ed. [34], below.
14. KATRAK, H., 'An empirical test of comparative cost theories: Japan, Peru, the United Kingdom, and the United States', *Economica*, 1969
15. —— 'An application of Lancaster's consumer demand theory to some recent hypotheses of international trade', A.U.T.E. Conference, 1973.
16. KEESING, D. B., 'Labor skills and international trade: evaluating many trade flows with a single measuring device', *Review of Economics and Statistics*, 1965.
17. —— 'The impact of research and development on United States trade', *Journal of Political Economy*, 1967.
18. —— Labour skills and the structure of trade in manufactures', in P. B. KENEN, and R. LAWRENCE, eds., *The Open Economy*, 1968.
19. —— 'Different countries' labour skill coefficients and the skill intensity of international trade flows', *Journal of International Economics*, 1971
20. KENEN, P. B., 'Nature, capital and trade', *Journal of Political Economy*, 1965.
21. —— 'Skills, human capital and comparative advantage', in W. LEE Hansen, ed., *Education, Income and Human Capital*, Univ. Nat. Bur. Comm. Econ. Res., Madison Wisc., 1968.
22. LANCASTER, K. J., 'A new approach to consumer theory', *Journal of Political Economy*, 1966
23. LEONTIEF, W. W., 'Factor proportions and the structure of American trade: further theoretical and empirical analysis', *Review of Economics and Statistics*, 1956
24. MARKHAM, J. W., 'Market structure, business conduct and innovation', *American Economic Review*, 1965.
25. OHLIN, B., *Interregional and International Trade*, revised ed., 1967.
26. POSNER, M. V., 'International trade and technical change', *Oxford Economic Papers*, 1961.
27. RAY, G. F., 'The diffusion of new technology: a study of ten processes in nine countries', *National Institute Economic Review*, 1969.

APPENDIX TABLE A IV

U.S./U.K. comparative exports to world markets

Industrial classification	*U.S./U.K. Exports*		
	1962	*1964*	*1966*
Aircraft	10.53	8.47	2.93
Motor vehicles and railway equipment	1.05	1.14	1.41
Electrical engineering	1.63	1.83	1.90
Scientific instruments	2.30	2.14	2.30
Chemical and allied products	1.92	1.98	2.00
Mechanical engineering	1.67	1.91	1.85
Rubber and rubber products	1.20	1.43	1.53
Glass, pottery, and building materials	1.14	1.15	1.21
Petroleum products	1.30	1.50	1.34
Metal products	1.08	1.15	1.40
Non-ferrous metals	0.98	1.20	1.01
Iron and steel	0.84	1.10	0.92
Clothing, footwear, and leather	0.83	0.90	1.00
Paper, printing, publishing	2.02	2.19	2.37
Food, drink, tobacco	1.39	1.27	1.05
Textiles	0.71	0.73	0.73
Timber and furniture	1.83	1.89	2.20

SOURCES O.E.C.D. Foreign Trade Statistical Bulletin, Series B, 1962, 1964, and 1966

APPENDIX TABLE A V

Skill intensities in United Kingdom industries

Industrial classification	*Large plants*	*Small plants*	*All plants*
Vehicles (a)	0.69	1.28	0.76
Engineering and electrical goods (b)	0.54	1.06	0.71
Chemicals and allied industries (c)	0.43	0.34	0.42
Rubber and plastic products	0.34	0.35	0.34
Glass, pottery, building materials	0.51	0.57	0.56
Manufacture of metal goods	0.65	0.72	0.43
Clothing, footwear, leather	1.62	2.55	2.50
Paper, printing, publishing	1.18	3.36	2.21
Food, drink, tobacco	0.22	0.31	0.27
Textiles	0.38	0.46	0.56
Timber and furniture	1.01	1.56	1.49

(a) Including aircraft

(b) Including mechanical engineering and scientific instruments.

(c) Including petroleum products

For notes and sources of data see Appendix notes.

REFERENCES

1. Baldwin, R. E., 'Determinants of the commodity structure of U S trade', *American Economic Review*, 1970.
2. Bhagwati, J. N., 'The pure theory of international trade. a survey', in *Trade, Tariffs and Growth*, 1969
3. ——— and Bharadwaj, R., 'Human capital and the pattern of foreign trade the Indian case', *Indian Economic Review*, 1967.
4. Comanor, W. S., and Scherer, F. M., 'Patent statistics as a measure of technical change', *Journal of Political Economy*, 1969
5. Findlay, R., *Trade and Specialisation*, 1970.
6. Gruber, W., Mehta, D , and Vernon, R , 'The R and D factor in international trade and investment of United States industries', *Journal of Political Economy*, 1967.
7. —— and Vernon, R , 'The technology factor in a world trade matrix', in R Vernon, ed [34], below
8. Harman, A , *The International Computer Industry*, 1971
9. Hirschman, A. O., *Development Projects Observed*, 1967.
10. Hufbauer, G. C., *Synthetic Materials and the Theory of International Trade*, 1965.
11. —— 'The impact of national characteristics and technology on the commodity composition of trade in manufactured goods', in R. Vernon, ed [34], below.
12. Johnson, H. G , *Comparative Cost and Commercial Policy Theory for a Developing World Economy*, Stockholm, 1968.
13. —— 'The state of theory in relation to the empirical analysis', in R Vernon, ed. [34], below.
14. Katrak, H , 'An empirical test of comparative cost theories Japan, Peru, the United Kingdom, and the United States', *Economica*, 1969.
15. —— 'An application of Lancaster's consumer demand theory to some recent hypotheses of international trade', A U T.E. Conference, 1973.
16. Keesing, D. B., 'Labor skills and international trade · evaluating many trade flows with a single measuring device', *Review of Economics and Statistics*, 1965.
17. —— 'The impact of research and development on United States trade', *Journal of Political Economy*, 1967.
18. —— 'Labour skills and the structure of trade in manufactures', in P. B. Kenen, and R. Lawrence, eds , *The Open Economy*, 1968.
19. —— 'Different countries' labour skill coefficients and the skill intensity of international trade flows', *Journal of International Economics*, 1971
20. Kenen, P. B., 'Nature, capital and trade', *Journal of Political Economy*, 1965.
21. —— 'Skills, human capital and comparative advantage', in W. Lee Hanson, ed , *Education, Income and Human Capital*, Univ. Nat. Bur. Comm. Econ Res., Madison Wisc , 1968.
22. Lancaster, K. J , 'A new approach to consumer theory', *Journal of Political Economy*, 1966
23. Leontief, W. W., 'Factor proportions and the structure of American trade: further theoretical and empirical analysis', *Review of Economics and Statistics*, 1956
24. Markham, J. W , 'Market structure, business conduct and innovation', *American Economic Review*, 1965.
25. Ohlin, B , *Interregional and International Trade*, revised ed , 1967
26. Posner, M. V., 'International trade and technical change', *Oxford Economic Papers*, 1961.
27. Ray, G. F., 'The diffusion of new technology: a study of ten processes in nine countries', *National Institute Economic Review*, 1969.

28. Report on the Census of Production 1963 Board of Trade, London, H.M.S O, 1968.
29. Statistics on Incomes, Prices, Employment and Production, Mar. 1965, Ministry of Labour, London, H.M.S.O, 1965.
30. Statistics on Science and Technology, Dept of Education and Science, Ministry of Technology, London, H.M S.O, 1967.
31. United States Census of Manufactures, 1963, Bureau of the Census, Washington, 1967.
32. United States Statistical Abstract, various years.
33. VERNON, R 'International investment and international trade in the product cycle', *Quarterly Journal of Economics*, 1966.
34. —— ed., *The Technology Factor in International Trade*, Univ. Nat. Bur. Comm. Econ Res, New York, 1970

THE INCOME ELASTICITY OF DEMAND FOR HOUSING[1]

By R K WILKINSON

I. Introduction

THERE exists an extensive literature on the demand for housing and the problems of estimating income elasticity of demand Apart from their academic interest, accurate estimates of income elasticity are of concern to policy makers and planners and to all who are involved in forecasting the demand for housing

Traditionally economists have tended to classify goods and services according to the value of the income elasticity of demand and there has been a tendency to regard housing as a 'necessity' in this technical sense. Early estimates of elasticity by both Engel and Schwabe gave coefficients below unity. In more recent times studies by Houthakker (1961), Leser (1961), and Lee (1964) also have produced estimates of coefficients of less than unity. This view of housing has been challenged in studies by Grebler (1952), Muth (1960), and Reid (1962) in the United States, and by Clark and Jones (1971) in the U K All these latter writers obtained coefficients greater than unity and, in addition, they all estimated the coefficient for owners to be significantly higher than that for renters The most recent study of U K data by Byatt, Holmans, and Laidler (1972) arrives at national estimates of between 0·7 and 0·9. A summary of the literature on this and related questions up to 1970 is to be found in de Leeuw (1971) and Wilkinson with Gulliver (1973)

The debate has been empirical in character centring principally on the statistical questions of the definition of variables, the specification of equations, and the technique of analysis Both the content of housing consumption and the selection of an appropriate definition of income are possible areas of disagreement Mortgage repayments, market values, rates, rents (both including and excluding repairs) have variously been used to measure housing expenditure Income measures either net or gross of taxation[2] and other stoppages relating to the household or the head of household on a weekly, monthly, or annual basis have similarly been used De Leeuw has

[1] This paper arises out of a study of House Prices and the Demand for Private Housing in Leeds supported by the S S R C from 1967 to 1969. I am pleased to acknowledge the assistance of Stuart Gulliver and Catherine Archer at different stages of this work

[2] There is a tendency to presume that taxation reduces incomes but in the U K at any rate the situation is rather more complex than this Middle-income house purchasers receive tax concessions which may sometimes amount to substantial subsidies According to the level of aggregation the omission of the effect of such concessions could result in an upward bias in the value of the elasticity coefficient

argued that the non-money income owner-occupiers receive from the rental values of their dwelling needs also to be included A much more important distinction, perhaps, is whether an attempt is made to exclude the transitory components of income and expenditure Reid and later Lee each have used a measure of 'permanent' income. The scope of individual investigations has varied from cities and parts of cities to nations, from homogeneous tenure groups to all consumers and from cross-sections of observations for single years to time series of varying length. In testing hypotheses some investigators have sought to take account of the influence of various socio-demographic variables such as family size, age, and socio-economic class of the head of the household. Lee, in particular, makes extensive use of socio-demographic variables the inclusion of which he argues reduces the chance of overestimating the income elasticity of demand

There are obviously many purely statistical reasons why measures of income elasticity might differ, from the single point of view of forecasting income elasticity, however, the precise value of the coefficient may be less important than its stability over time Measures which are unrelated to a behavioural model, however, have obvious disadvantages The main concern of this paper therefore is with the conceptual problems involved in the explanation and understanding of behaviour Thus the first part is concerned with clarifying some problems of measurement and with the interpretation of results

II. Influences on the demand for housing

The definition of housing

One of the interesting features of the debate on the value of income elasticity lies in the implicit view that housing, to use the terminology still favoured by some textbooks, is *either* a 'necessity' or a luxury' It is difficult to find examples of any goods and services which are of homogeneous quality and which in the consumer's eyes possess a single attribute which satisfies a simple demand Man's ingenuity has usually managed to discover a basis for differentiating products and services as a means of charging a higher price whilst technical progress plus the widespread application of 'marketing' techniques have combined to enhance differences (both real and apparent) in the quality of goods and the attributes they possess It is misleading therefore to regard most products as possessing a single use and a single quality, and housing is no exception. Housing is perhaps most accurately viewed as a set of (potential) 'services' available to satisfy a complex set of household demands which range from the basic need for shelter, through the provision of sanitation and other services to the

possession of a favoured location.[1] Each of these sets of services may be provided at different levels of quality and the number of services available itself is an index of housing quality. Thus an individual's dwelling-place may reflect (like perhaps his car or his clothes) not only his attempt to satisfy a basic need but also his social and economic status and aspirations.

The expenditure function

The definition of housing quality will in turn tend to identify the characteristic purchaser in terms of his socio-economic attributes. The pre-occupation of the studies cited above with both the precise measure of housing consumption and the more complete specification of the demand function suggests an implicit recognition of this point. David (1961) and Clark and Jones (1971) have explored the relationship between expenditure and family size; Lee (1964) and de Leeuw (1971) have sought to hold these influences constant in estimating income elasticities. No unifying model of household behaviour with regard to the consumption of housing exists, however, either as the basis for or as the result of these studies and this necessarily widens the scope for the interpretation of estimates of elasticity. Thus expenditure on housing can be seen as the outcome of three sets of influences on consumers, viz. their needs, their aspirations, and their ability to realize their needs and aspirations. The latter is represented by a measure of income and the two former may be regarded as constraining or qualifying the way in which income and changes in income affect housing expenditure. All three sets of influences might be expected to vary with the stage of the life cycle of the household as well as with its size and composition.

In formal terms the expenditure function of an individual household i ($i = 1, ..., n$) in a group of n households at period t may be represented as follows:

$$h_{it} = a_i y_{it} + b_i x_{it} + c_i z_{it} + e_{it} \qquad (1)$$

where y denotes a measure of disposable income, x is a measure (or measures) of 'need', z is a measure (or measures) of aspiration, a_i, b_i, and c_i are co-efficients for household i, and e is an error term. It is possible but perhaps not realistic to show each set of influences as determined in separate equations since the measures interact. For example, housing aspirations might be represented broadly by the form of tenure and perhaps more narrowly by social class, occupation, and income group. The basic physical require-

[1] One can in fact postulate the existence of a hierarchy of 'housing services'. Individual attributes of the dwelling compared with the environment may be grouped together (see Wilkinson (1971)) and these in turn may be divided into sub-groups corresponding to the broad services provided by the dwelling very much after the utility-tree approach of Strotz (1959) and Gorman (1959).

ments of the household similarly might be defined in terms of its size and composition The latter provides some problems of measurement but the age of the head could be taken as providing a rough indication of the stage of the life cycle reached by the household The ability to purchase housing is fundamentally determined by the level of income but, as already indicated this is not independent of the other variables listed which may act as (maximum and minimum) limits at different stages in the household's life cycle

The structure of housing expenditure

In order to consider the likely effects of increases in any of the independent variables it is convenient to distinguish between physical consumption and spending This is also one way in which we can try to distinguish quality effects on spending Thus an increase in income, all other things being equal, may result in higher expenditure on housing but no increase in the number of housing services, i e higher quality housing is purchased Increasing family size may affect both real consumption and money expenditure depending partly on whether there are any economies of scale to be had for the household from the current dwelling It is conceivable that an increase in physical consumption could be had at the expense of the quality of housing with the level of expenditure remaining unchanged Equally, following the view that the dwelling is fundamentally a collection of housing attributes or potential services, a householder may opt for a different bundle of services within a given price range Rising aspirations are more likely, *ceteris paribus*, to give rise to increases in expenditure than physical consumption through perhaps changes in tenure if not in the quality of housing Thus there exists a variety of possible behaviour patterns each of which may give rise to elasticity coefficients of differing values In considering consumption rather than the expenditure the relative prices of the housing services become relevant and need to be specified as parameters of the system If these are included we can postulate a demand function for each attribute or service provided by a dwelling for which there is a corresponding utility function

If we postulate a household utility function, the reactions of housing expenditure to changes in any of the parameters of the function will be governed by the elasticity of substitution between housing expenditure and all other expenditures and within the housing category between essential and non-essential housing services It seems plausible to assume a constant elasticity of substitution for housing compared with non-housing and also between individual and sets of 'housing services'. If this is the case it can be shown that the marginal utility of any single (set of) housing service(s) is positive and diminishing and that for the substitute goods

comprising the function, the marginal utility of one increases with the other [1]

In order to aggregate household demand function into market functions we can make the simplifying assumptions that the consumption function for each household is linear and that housing consumption is simply a function of income Following the notation used in equation (1) we may write

$$h_{it} = a_i y_{it} + b_i + e_i \tag{2}$$

An aggregation of households over any sector (defined in terms of area or tenure for example) or for the whole economy gives

$$H_{it} = \sum_{n=1}^{n} h_{it}, \qquad Y_{it} = \sum_{i=1}^{n} y_{it},$$

where the capital letters denote market as opposed to household variates Similarly,

$$b_t = \sum_{i=1}^{n} b_i \quad \text{and} \quad e_t = \sum_{i=1}^{n} e_i$$

Thus aggregate expenditure on housing depends on the incomes of the n households plus an error term whose expected value is zero for all the y_{it}

Some statistical problems

As was pointed out above there are many problems connected with the specification of relationships and the measurement of variables and these have been dealt with fairly thoroughly in the literature (See especially de Leeuw, Lee, and Byatt, Holmans and Laidler) The fundamental questions are how far do estimates of a and b reflect the true relationship between consumption and income? And to what extent are they valid representations of household behaviour? If the equation is fitted by the least squares technique and all households have the same propensity to consume housing then the expected values of estimates of the constants in the regression equation will be valid estimates of the true values

Thus

$$H = \hat{a}Y + \hat{b} \tag{3}$$

where $\hat{a} = \sum_{i=1}^{n} a_i e_a$ and $\hat{b} = \sum_{i=1}^{n} b_i e_b$ The es are estimation errors

If all households have the same a_i and b_i, the sum of the e_a is unity and the sum of the e_b is zero and

$$E(\hat{a}) = a, \qquad E(\hat{b}) = b.$$

Similarly, if household incomes are related to total incomes in simple fashion by an equation

$$Y_{it} = a_i Y_t + B + e_{it} \tag{4}$$

[1] By conventional techniques we can show that both the individual demand for housing compared with other goods and services or the demand for a single housing service compared with other housing services is a diminishing function of its cost to the consumer See Appendix I

then by definition (of Y_t), $\Sigma x_i = 1$ and $\Sigma\beta = 0$ and $\Sigma e_{it} = 0$, and the aggregate equation validly represents the individual household behaviour. These two highly restrictive conditions, however, draw attention to the possible sources of bias in the aggregation of households, viz. variations in the propensities to consume housing among households and the variation of household income in relation to the average over a period. Unless all sub-groups (in terms of income and consumption) are represented in their proper proportions their biases may occur as a result of under- or over-representation. Thus estimates of income elasticity may vary as a consequence of the way in which sub-groups of households are aggregated to obtain expenditure functions. These problems of aggregation bias are well known (see Theil, 1971) and will not be discussed further here.

A related and important question concerns the behaviour of particular socio-economic sub-groups of households. If different socio-economic groups purchase differing amounts, compositions, and qualities of housing (as they do), we should expect their individual views on housing to differ and their behaviour, as represented by the coefficient of income elasticity, to vary accordingly. On the whole, owner occupiers belong to higher socio-economic groups than renters with incomes higher on the average though (as a consequence of age) spread over a wider range. Similarly, the housing they buy is regarded as being of generally higher if more variable quality. Thus, we would expect the income elasticities of owner occupiers to be higher than renters. It follows that in estimating income elasticities, if the quality of housing is not allowed to vary except within narrow limits and the effects of family size and composition are similarly held constant, the size of the coefficient of income elasticity is likely to be reduced.[1] The coefficient therefore will vary with both the precision of the definition of the socio-economic group and with the broadness of the definition of housing.

We may conclude that the close specification of the quality of housing services is likely to result in a *reduction* in estimates of income elasticity and that the narrow specification of the socio-economic class is likely to have the same effect. In so far as we can recognize the existence of groups of consumers who are relatively homogeneous both in their attributes and in their housing consumption then it is probably unrealistic to combine or 'average' their elasticities and to think in terms of a single elasticity value for all consumers.

[1] The statistical effect of holding constant the influence of individual parameters of the demand function which are correlated with expenditure will be to reduce the size of the regression coefficient. Thus, for a given level of income and expenditure, if the regression coefficient is reduced, it follows by definition that $\partial H/\partial Y$ (the 'marginal propensity to consume housing') is reduced. Since we can define the income elasticity of demand for housing in a linear and homogeneous function as *MPC* 1/*APC* (i.e. $\partial H/\partial Y \; Y/H$) it follows that the coefficient of income elasticity must also be reduced.

III. A study of a local housing market

Estimates of the income elasticity of demand for housing were made by the author as part of a study of owner-occupied housing in Leeds A detailed description of the data on which this and other studies were based is available in Wilkinson (1971) In brief, complete surveys were carried out of the mortgages granted by eleven building societies and the local authority for the years 1960 and 1964 for properties located in the area administered by Leeds C B C All the information contained in the mortgage proposal form was recorded (except the mortgagee's name) and this was supplemented by other data supplied by the Leeds Town Planning Department. This study of housing demand differs from others in this field in three main aspects First, it is concerned with a cross-section of purchasers (for owner occupation) taken at two points in time, second it makes use of alternative measures of income and consumption, third, it attempts to incorporate measures of housing quality. Housing is taken as a complex of 'services' internal and external to the dwelling as explained above The locational factor affecting the value of the dwelling was estimated by means of factor analysis (Wilkinson, 1971 and 1973) and this was taken as the most significant index of housing quality Though an alternative measure, viz distance of dwellings from the centre of the city, was also used as a proxy for quality [1]

The data also permitted experimentation with the age, occupation, and family size of the purchasers. The latter was of special significance since it enabled the testing of hypotheses suggested in the work of David (1961) Occupation was found to be highly and positively correlated with income and thus to save space, only the results for age of the head of household and family size are reported below Three measures of income and six measures of expenditure or consumption were analysed

Alternative measures of income and housing expenditure

The measures of income have been discussed in previous papers and concern basic income of the head of household (y_1), the total income of the head of household (y_2), and the total household income (y_3) It might be argued that, from the point of view of discussing the permanent income approach, y_1 is the most appropriate variable It is certainly the one which tends to figure most prominently when building societies are making loans (indeed it might be argued that building societies are interested in permanent income *per se*) Although there is some tendency for the coefficients of y_1 to be smaller than those for the other measures of income, this is not consistent for all the results In fact the correlation between y_1, y_2, and y_3 is so high

[1] Leeds is a city with a fairly pronounced concentric development and, in general, the quality of dwellings increases with distance from the centre

that the differences between coefficients are not statistically significant. Ideally, measures of housing consumption would take into account the market rent and the amount spent on rates and repairs. Unfortunately, we do not have measures of these latter, though it might be argued that the price of the dwelling will automatically reflect each of these variables. Most of the studies which have been reported in this field used selling-price as the dependent variable. Total repayments will differ according to the previous capital assets of the household in that the size of repayments is related to the size of the loan. These latter are likely to be related to the life cycle of the household, so that one may expect the coefficients relating repayments to income to be influenced more strongly by measures of household size and composition.[1] The use of price per square foot or price per room automatically introduces a quality dimension into the analysis. Measures of area in square feet are likely to be more sensitive as measures of quality than the number of rooms, first, because the latter can be measured only in discrete units; and, second, because it is quite likely that dwellings with the same number of rooms will have different areas.

These data were analysed on the basis of individual observations and by 'neighbourhood units'. In the sample years there were fifty-three of these latter which in simple terms are fairly self-contained and easily recognizable districts of the city. Averages of observations grouped on the basis of neighbourhood units were used as proxies for the permanent income and expenditure of the households moving in. This therefore recognizes the permanent component of income and expenditure as an expected value which can be represented by the arithmetic mean whilst the transitory component is represented by the variation around the mean. The tables below show estimates of income elasticity on the basis of individual and grouped observations for each of the sample years. The effect of specifying the function more precisely by taking account of family size and the locational qualities of the dwelling is also demonstrated.

Table I shows that the effect of using grouped observations is to raise the value of income elasticity in all cases. This result is consistent with those of other investigations of income elasticity of demand using permanent income. It shows that people are relatively sensitive to changes in their expected level of income whereas this is not true of all variations as in their actual income. The explanatory power of income as measured by the value of R^2 was substantial. With the exception of price per square foot it

[1] Since the proportion of income rule operated by building societies automatically results in an income elasticity of unity for a sample of *maximum* borrowers, mortgage repayments are potentially a source of bias in estimates of the elasticity of demand for housing. This is not true of measures of market price, however, or of price per square foot or per room all of which for the households covered in this study are highly (positively) correlated with each other.

explained between 24 and 34 per cent of the variations in the dependent variables in each year for individual observations; in the case of price per square foot it explained only 9 per cent For grouped observations R^2 was very high; the lowest values were obtained for price per square foot (viz. 37 and 41 per cent respectively in 1960 and 1964); the other values were of the order of 60–70 per cent.[1] Price was expressed as a ratio of the area and the

TABLE I

Estimates of income elasticity of demand for housing 1960 and 1964

Dependent variable	*Individual observations*	*Grouped observations*
Price	0 684	1 258
	(0 037)	(0 138)
	0 806	1 525
	(0 034)	(0 122)
Price per room	0 398	0 855
	(0 033)	(0·120)
	0 582	1 141
	(0 030)	(0 140)
Price per square foot	0 351	0 767
	(0 051)	(0 143)
	n a.	0 900
		(n a)
Repayments	0 616	0 913
	(0 040)	(0 092)
	0 670	1 060
	(0 028)	(0 090)

NOTES:

(i) The 1964 values are shown below those for 1960.
(ii) The variables were fitted in natural logarithms.
(iii) Standard errors are shown in parentheses.

number of rooms because it was thought that these would provide alternative measures of size or quality. In fact the elasticity coefficients obtained are not significantly different and therefore only the results for price per room are reported below.[2]

The coefficients do differ significantly, however, according to the measure of the dependent variable. In particular the introduction of the quality dimension produces a smaller coefficient of elasticity in all but one case. The use of repayments compared with price also gives rise to slightly lower

[1] This is to be expected from grouped observations See Cramer (1964).

[2] A further reason for this choice was that the number of rooms (i e how space is used) would appear to be more significant from the consumer point of view than the space itself. In so far as area and number of rooms are not perfectly correlated, it is the number of rooms which is likely to be the most significant factor Further, if it is correct to suppose that price per square foot reflects rates and repairs by taking into account dwelling size we might infer that a more precise specification of the consumption function is necessary where rate and repair payments are specified as independent variables.

values of elasticity indicative of a slight downward bias which is presumably a consequence of some borrowers borrowing considerably below their maximum and for shorter periods. A correction was made to take account of the latter (by including the length of loan as the dependent variable) and this did raise the values of the coefficient slightly though not significantly

TABLE II

Estimates of partial income elasticity, 1960 and 1964

(All observations)

Dependent variable	b_{12}	$b_{12\cdot3}$	$b_{12\cdot4}$	$b_{12\cdot34}$	$R^2_{1\cdot234}$ %
Price	0 684	0 707	0 467	0 481	62 99
	(0 037)	(0 037)	(0 029)	(0 030)	
	0 806	0 809	0 446	0 432	60 59
	(0 034)	(0 035)	(0 029)	(0 029)	
Price per room	0 398	0 414	0 192	0 199	56 56
	(0 033)	(0 033)	(0 026)	(0 026)	
	0 582	0 592	0 285	0 286	53 11
	(0 030)	(0 030)	(0 026)	(0 026)	
Repayments	0 616	0 647	0 454	0 484	42 46
	(0 040)	(0 040)	(0 038)	(0 038)	
	0 670	0 671	0 477	0 468	43 61
	(0 028)	(0 029)	(0 028)	(0 029)	

NOTES

(i) Subscripts refer to the following variables.

1 The dependent variable ($\log_e$)

2. Income of head of household ($\log_e$)

3. Family size

4 Location factor

Thus b_{12} shows the 'gross' income elasticity, $b_{12\cdot3}$ shows income elasticity with the effect of family size held constant, $b_{12\cdot4}$ shows the effect of holding 'location' constant and $b_{12\cdot34}$ shows the effect of holding both family size and location constant

(ii) The results for 1960 are shown above those for 1964 in each case. Standard errors are shown in parentheses

The effect of taking account of the size of families and the locational attributes of dwellings is shown in Table II In each case the effect of family size alone is negligible but it raises the value of the coefficient whereas the location variable reduces it The combined effect is to reduce the size of the coefficients. These results are consistent with expectations in that the more precise specification of housing consumption (by holding constant the effect of location) leads to a reduction in value of the income elasticity For price per room it appears to be in the order of 0·2 to 0·3

The pattern of results in Table III exhibits broadly the same tendencies the value of the coefficient is reduced by a narrower specification of the services consumed. In both tables the results for 1964 are higher than those

for 1960. It is possible that this reflects either a trend in the income consumption relationship or bias in the samples though in some cases the differences are within the bounds of statistical error. Recalculations on the basis of standardized samples has not provided evidence of bias and we infer therefore that the differences may reflect the somewhat easier

TABLE III

Estimates of partial income elasticities, 1960 and 1964

(Grouped observations)

Dependent variables	b_{12}	$b_{12.3}$	$b_{12.4}$	$b_{12.34}$	$b_{1.234}$%
Price	1 258 (0 138)	1 061 (0 153)	0 888 (0 152)	0 701 (0 156)	72 35
	1 525 (0 140)	1 526 (0 140)	1 110 (0 107)	1 103 (0 108)	86 97
Price per room	0 855 (0 120)	0 707 (0 134)	0 493 (0 125)	0 353 (0 130)	70 96
	1 141 (0 122)	1 144 (0 120)	0 772 (0 091)	0 780 (0 192)	84 80
Repayments	0 913 (0 092)	0 957 (0 107)	0 719 (0 107)	0 667 (0 118)	72 62
	1 060 (0 090)	1 061 (0 090)	0 804 (0 072)	0 801 (0 073)	87 09

NOTE.

(i) 1964 values are shown below 1960 values in each cell

(ii) Subscripts and the interpretation of the *b*s are as in Table II. (See note (i).) Standard errors are shown in parentheses

credit facilities available in 1964 plus the entry of the local authority into the mortgage market The variable used to measure the influence of location in Table III differed from that in Table II It was not possible to derive an index by means of a factor analysis and therefore distance from the centre of the neighbourhood unit to the central business district was used as a proxy In a city like Leeds which is fairly concentric in its form and in which the quality of the neighbourhood rises with distance from the centre, such a measure is in fact a good proxy

Although the size of family does not appear to exert a very strong effect on the housing income relationship (and this is consistent with the findings of David (1961) and Reid (1962), mentioned above) it is possible that the results might differ for individual strata of consumers as a consequence of the stage of the 'life cycle' of the household discussed above The best indicators of 'life cycle' which were available for analysis were 'age of head of household' and the 'number of of children' The results of analysing 'repayments' (corrected for length of load) by these two variables is shown in Table IV.

These results show some tendency for the elasticity coefficient to fall over the lifetime of a childless couple and to rise slightly for all age groups up to four children and to fall for five and six children. An examination of the table confirms these tendencies, on the whole income elasticity tends to be higher for those with one to four than it does for those with no children or five to six children. Elasticity tends to increase with the age of head up to 55 years.

TABLE IV

Estimates of income elasticity for given types of family composition

(All observations)

Number of children	*Median age of head (years)*				
	23 5	29 5	39 5	49 5	59 5
0	1 135	0 630	0 951	0 647	.
	(0 168)	(0 077)	(0 149)	(0 177)	
	0 919	0 503	0 844	0 496	0 605
	(0 101)	(0 053)	(0 127)	(0 132)	(0 161)
1 and 2		0 855	0 886	1 005	0 692
		(0 090)	(0 087)	(0·154)	(0 249)
		0 781	0 729	0 612	0 710
		(0 060)	(0 059)	(0 111)	(0 177)
3 and 4	.	0 885	1 010	1 108	
		(0 190)	(0 088)	(0 449)	
		0 743	0 847	0 949	
		(0 137)	(0 075)	(0 374)	
5 and 6		1 013	0 561	0 654	..
		(0 384)	(0 109)	(0 265)	
		0 861	0 513	0 419	0 533
		(0 269)	(0 091)	(0 191)	(0 140)

NOTE.
Standard errors are shown in parentheses

The values of estimates for the highest age group are unreliable since they are based on relatively small samples (under twenty); the sample sizes for the other age groups are of the order of 200 observations. Although age and income are likely to be correlated, these results are compatible with a family life cycle which in the early stages is characterized by roughly one-for-one changes in housing expenditure and income but as aspirations are achieved this ratio falls The effect of increasing family size is to raise or maintain the income elasticity. It is possible that tastes for housing will vary with the age of the head of household and it is obvious that physical requirements will vary with the number (irrespective of the sex) of children There is no evidence available to assist in the analysis of the former but it is clear that the latter may well account for the low values of elasticities

bserved for the large families, i.e. it is physical requirements which are aramount for all except the highest income group In this case the results uggest the existence of economies of scale in the demand for housing If here are economies, one would expect the income elasticity to be stable over range of family size and then to decrease in the highest ranges (assuming heir incomes and aspiration level are taken as given) This result is uite compatible with the form of utility function specified above

Finally, we turn to the analysis of two measures of real consumption viz total area covered by the dwelling and its number of rooms) contained n Tables Va and b. Price was included as an explanatory variable at this tage and thus estimates of price elasticity are obtained

TABLE Va

Estimates of income elasticity for alternative measures of real consumption

(All observations)

Dependent variable	b_{12}	$b_{12\,3}$	$b_{12\,34}$
Total area	0·381	0·496	0·480
	(0·032)	(0·019)	(0·031)
	0·303	0·399	0·351
	(0·022)	(0·022)	(0·024)
Total rooms	0·282	0·312	0·293
	(0·020)	(0·021)	(0·023)
	0·225	0·218	0·184
	(0·015)	(0·016)	(0·018)

TABLE Vb

(Grouped observations)

Dependent variable	b_{12}	$b_{12\,3}$	$b_{12\,34}$	$b_{12\,345}$
Total area	0·509	0·679	0·610	0·593
	(0·090)	(0·091)	(0·098)	(0·104)
	0·524	0·813	0·776	0·816
	(0·092)	(0·098)	(0·101)	(0·091)
Total rooms	0·411	0·329	0·339	
	(0·034)	(0·035)	(0·033)	
	0·383	0·302	0·364	0·334
	(0·038)	(0·042)	(0·015)	(0·040)

NOTES·

(i) Results for 1960 shown above those for 1964 in each cell

(ii) Independent variables are as follows

2. Income of head of household,
3 Price of dwelling,
4. Family size,
5. Location (distance from centre),

The *b*s are interpreted as in Table II, Note (i)

The coefficients of income elasticity exhibit the same tendencies as thos(described for expenditure Coefficients relating to neighbourhood unit: are substantially larger than those obtained for all observations. Estimate: of the partial coefficients are arrived at after including price as an indepen dent variable In the case of total area, this causes the size of the coefficien to rise, whereas for price per room it causes it to fall

TABLE VI

Estimates of the elasticity of demand of dwelling area and number of rooms t(price, income, and family size

(Grouped observations)

Dependent variable		*Price*	*Income*	*Family size*	*Location*	R^2
Total area		− 0 237	0 679		.	0 555
		(0 079)	(0 091)			
	1960	−0 232	0 610	0 153		0 581
		(0 071)	(0 098)	(0 090)		
		−0 272	0 593	0 164	0 038	0 584
		(0 111)	(0 104)	(0 094)	(0 076)	
		−0 291	0 813			0 590
		(0 064)	(0 098)			
	1964	−0 244	0 776	0 183		0 597
		(0 072)	(0 101)	(0 132)		
		−0 567	0 816	0 176	0 226	0 687
		(0 096)	(0 091)	(0 118)	(0 061)	
Total rooms		0 195	0 329	.	.	0 810
		(0 029)	(0 035)			
	1960	0 214	0 339	0 008		0 859
		(0 026)	(0 033)	(0 038)		
	1964	0 166	0 302	.		0 674
		(0 029)	(0 042)			
		0 250	0 364	0 249		0 723
		(0 011)	(0 015)	(0 024)		

Table VI shows that our estimates of price elasticity have a negative sig in the case of area and a positive sign in the case of rooms This suggest that as price rises, consumers tend to buy smaller houses (or, what amount to the same thing, to spend the same proportion of income on more expensiv houses) The positive correlation between number of rooms and pric suggests that the number of rooms is not really an appropriate measure o consumption, and that price is really the appropriate dependent variabl In other words, the number of rooms purchased is more likely to be a func tion of family size or income The addition of location into the analysi tends, as before, to reduce the size of the coefficient of income elasticity I general, coefficients for real consumption are substantially lower than thos

for expenditure This is compatible with the view discussed above that the demand for space is relatively inelastic, whereas the demand for quality is rather more elastic Thus a household's real requirements tend to be inelastic with respect to income, but spending on housing tends to increase at higher levels of income implying that housing of a higher quality is bought

IV. Summary and conclusions

It has been argued that estimates of the elasticity of demand for housing are likely to vary according to the definition of the services included as housing, the characteristics of the purchaser, and the extent of the housing market Thus we should expect to obtain different estimates according to the tenure group under consideration, the socio-demographic attributes of the purchasers, and the quality of the housing Equally, in so far as these factors vary in their geographical distribution, we should expect regional variations around the national (average) estimates [1] Although it is apparent from a brief survey of the literature that a wide range of estimates of income elasticity have been reported it would appear possible to reconcile these estimates within the analytical framework which has been set up

Different but consistent estimates were obtained according to how housing consumption was measured (i e in terms of price, price per square foot or per room, repayments) whether it was measured in real or value terms and whether or not the locational features of the dwelling were held constant Variations arising from the use of different income concepts were relatively small and therefore measurements on the basis of head of household income only are reported. The effect of family size on estimates for the whole sample were slight but a disaggregation by age of the head of household and the number of children showed that these aspects of family composition did exert an effect on the elasticity not incompatible with expectations of the differing aspirations, needs, and possibilities at different stages of the life cycle of a household and also with the existence of economies of scale This latter conclusion gains some support from the low sensitivity of both dwelling area and the number of rooms to family size

The emphasis of this paper has been on the analysis and comprehension of housing behaviour but, as was pointed out at the beginning, an accurate measure of elasticity is also important for forecasting purposes The twin objectives of explanation and prediction are not necessarily in conflict though no decisive answer has been offered (or attempted) to the question

[1] A recognition of this point is contained in Reid (1962), p 348 'The coefficient (of income elasticity of demand for *rooms*) appears to be around 0 4 This is appreciably less than coefficients of housing consumption with respect to normal income of around 2 0 It has its counterpart in the demand for calories and for total food ' This is a point which subsequent investigators have missed or chosen to ignore

of 'What is *the* elasticity of demand for housing?' because this is not a question which can be answered unless the relevant socio-economic group (and by implication its type of housing) is specified. It is interesting in considering the whole range of results which have been produced over time to see that for the narrower definitions of housing the variation is relatively small; variation increases as housing is taken to include a wider selection of services. This supports the view that there are 'necessary' components in housing demand common to most consumers but that for higher standards of living, variations in tastes and habits become more significant and variations in the value of coefficients of elasticity are to be expected

The implication of this paper for forecasters is that it is important to be precise and consistent in the specification of the demand function and in the measurement of variables Various measures of elasticity are possible and some will be more appropriate for a given purpose than others· the question of the selection of an appropriate measure will be resolved partly by the availability of data and partly by the purpose in hand.

University of Sheffield

REFERENCES

BYATT, I C. R, HOLMANS, A E., and LAIDLER, D. E. W., 'Income and the demand for housing. some evidence for Great Britain', *Economic Notes No. 1 Directorate General of Economics and Resources, Department of the Environment*, 1972.

CLARK, COLIN, and JONES, J T., 'The demand for housing', University Working Paper, *Centre for Environmental Studies*, 1971.

CRAMER, J. S, 'Efficient grouping, regression and correlation in Engel curve analysis', *Journal of the American Statistical Association*, 59 (1964), pp 233-50.

DAVID, M H., *Family Composition and Consumption*, North Holland Publishing Company, 1961

DE LEEUW, F, 'The demand for housing a review of the cross section evidence', *The Review of Economics and Statistics*, vol 53, no 1, 1971.

GORMAN, W M, 'Separable utility and aggregation', *Econometrica*, vol. 27, no. 3, 1959

GREBLER, L. T., *Housing Market Behaviour in a Declining Area*. Columbia, 1952

HOUTHAKKER, H. S, 'The present state of consumption theory', A Survey Article, *Econometrica*, vol 29, no. 4, 1961.

LEE, T. H., 'The stock demand elasticities for non-farm housing, *The Review of Economics and Statistics*, vol. xlv, no. 2, 1964.

—— 'Housing and permanent income. texts based on a three year reinterview survey', *The Review of Economics and Statistics*, vol. xlxi, no. 2, 1968.

LESER, C. E. V, 'Commodity group expenditure functions for the United Kingdom, 1948–57', *Econometrica*, vol. 29, pp. 24–32, 1961.

MUTH, R. F., 'The demand for non-farm housing', in A. C. Harberger, *The Demand for Durable Goods*, pp. 29–96, 1960.

REID, M. G, *Housing and Income*. University of Chicago Press, 1962.

STROTZ, R. H., 'The empirical implications of a utility tree', *Econometrica*, vol 25, no. 2, 1959.

THEIL, H., *Principles of Econometrics*, North Holland Publishing Company, 1971

WILKINSON, R. K., 'The determinants of house prices', Centre for Environmental Studies, Conference Papers, 1971.

WILKINSON R. K., 'House prices and the measurement of externalities', *Economic Journal*, vol. 83, no. 329, 1973.

—— and ARCHER, C. A., 'Measuring the determinants of house prices', *Environment and Planning*, vol. 5, pp. 357–67, 1973.

—— and GULLIVER, S., 'The economics of housing a survey', *Social and Economic Administration*, vol. 5, no 2, 1971 Revised and reprinted in *Survey of Recent Developments in Social Policy and Social Administration*, edited by M. H Cooper, Heinemann, 1973.

APPENDIX

The application of the C.E.S utility function to housing

Let us assume that the utility function for both housing (h_1) and non-housing (h_2) and essential (h_1) and 'less essential' housing (h_2) is linear and homogeneous with an elasticity of substitution (c) which is constant. The following theorem applies to both cases

$$U = M(Ah_1^{-\rho} + Bh_2^{-\rho})^{-(1/\rho)} \tag{1}$$

where $A > 0$, $B > 0$, $M > 0$, $-1 < \rho < 0$ or $0 < \rho$

The constraints imposed are $h_1 \geqslant 0$, $h_2 \geqslant 0$, $U \geqslant 0$

Take all first and second partial derivatives

$$\frac{\partial U}{\partial h_1} = AM^{-\rho}\left(\frac{U}{h_1}\right)^{1+\rho} \tag{2}$$

$$\frac{\partial U}{\partial h_2} = BM^{-\rho}\left(\frac{U}{h_2}\right)^{1+\rho} \tag{3}$$

$$\frac{\partial^2 U}{\partial h_1^2} = AM^{-\rho}(1+\rho)\left(\frac{U}{h_1}\right)^{\rho}\frac{(\partial U/\partial h_1)h_1 - U}{h_1^2} \tag{4}$$

$$\frac{\partial^2 U}{\partial h_2^2} = BM^{-\rho}(1+\rho)\left(\frac{U}{h_2}\right)^{\rho}\frac{(\partial U/\partial h_2)h_2 - U}{h_2^2} \tag{5}$$

$$\frac{\partial^2 U}{\partial h_1 \partial h_2} = AB(1+\rho)M^{-2\rho}U^{1+2\rho}(h_1 h_2)^{-(1+\rho)} \tag{6}$$

For positive M, h_1 and h_2, A and B must be positive for (3) and (4) to be positive

From Euler's theorem

$$U = \frac{\partial U}{\partial h_1}h_1 + \frac{\partial U}{\partial h_2}h_2. \tag{7}$$

Equations (2), (3), and (7) show that the numerators of the right-hand side of (4) and (5) are negative, thus (4) and (5) are opposite in sign to $1+\rho$ and will be negative only if $\rho > -1$ which is assumed above. Thus for positive A and B the marginal utility of h_1 is positive and diminishes as its level rises Since the derivatives of (6) are positive it follows that the marginal utility of h_1 increases with h_2.

RATES OF RETURN TO PHYSICAL CAPITAL I MANUFACTURING INDUSTRIES IN ARGENTINA[1]

By AMALIO HUMBERTO PETREI

I. Data and methods

1 *Introduction*

THE purpose of this paper is to estimate the rates of return to physic capital in manufacturing industries in Argentina for each of the yea 1961–7. Physical capital is here defined to include both fixed assets a inventories

The study covers most of the industrial (manufacturing) groups und the International Standard Industrial Classification at the three-digit lev Given the disparity in the size of groups and the way in which the samp was drawn, the results are presented for thirty-seven industrial groups

The estimates have been done for the corporate sector only. Of cours it would have been desirable to work with the noncorporate sector as we but the necessary data were not available.

Estimates of rate of return are useful for several purposes:

1. They provide a basic tool for project evaluation, both social a private The discussion about the appropriate discount rate for soci projects has been thoroughly reviewed in the literature. Professor Ha berger has concluded that the appropriate rate is a weighted average of t private sector marginal productivity of capital and the net of tax yield c private savings The estimates obtained in Section II provide usef information to approximate that rate.

2 Knowledge of rates of return in different sectors would also allow to evaluate whether the allocation of capital among different areas efficient or not. If we can show that there are significant differences amo rates of return of different sectors, that these do not tend to narrow dow over time, and that investment in the sectors with higher rates of retu does not increase in relation to other sectors, generally this would const tute evidence of a malfunctioning of the capital market A partial evalu tion along this line is made in Section II.

[1] This article is based on material included in the Ph D dissertation presented in t Department of Economics of the University of Chicago. The author would like to tha Professors Arnold C Harberger, H Gregg Lewis, and Larry Sjaastad for helpful suggestio received during the preparation Nevertheless, all mistakes and shortcomings are, of cour the author's responsibility

An important part of this research was done at the University of Cordoba (Argentin and completed at the University of Chicago. The investigation was partially financed b two grants received from Consejo Nacional de Investigaciones Científicas y Técnicas Argentina

2. *Sources of data, list of corporations, and classification by industry*

The principal source of data consists of corporation balance sheets, which Argentine corporations must regularly submit to the 'Inspección General de Justicia' in the district in which they are registered

Given the resources at my disposal and the method to be followed for calculating rates of return, I could not hope to make such calculations for all corporations I decided therefore to select a stratified random sample from a list constructed on the basis of information for 1963 The list comprised 3,831 corporations.

The year 1963 was chosen as the year of sample selection because I considered it one of the best suited for my purposes. It is close to the middle of the period (1961–7) under consideration, and I could be sure that re-evaluation of assets had been made in the largest number of cases But besides that, the most important consideration was that in 1964 an economic census was taken in Argentina based on 1963 data. This would permit comparisons to be made and the information and results to be extended to the national level.

For purposes of preparing the list the registration number of the corporation, its main product category, and total sales were taken for 1963 Once I had the list I proceeded to classify the firms by primary product according to the I S I C (International Standard Industrial Classification) (as of 1958) at the three-digit level. The main product category was taken from the 1963 balance sheet Since some of the corporations produced secondary products that would fall under different I S I C numbers (industry) than that of their primary products, the classification of the enterprises by industry is somewhat ambiguous For example, I know that the sample corporations I have classified in the beer industry also have substantial sales of soft-drink products However, in general the balance-sheet data even on sales are not recorded by detailed product Thus I cannot estimate how much ambiguity (overlapping of industries) is produced by the classification of the firms by their *principal* product

For the purpose of designing the sample, I was interested in having a good proxy variable for physical capital Book values as recorded cannot be used for this purpose without adjustment because Argentina has experienced a long history of inflation at variable rates So I worked out simple rough estimates of physical capital that were intended to be used only for stratification purposes

On the basis of the information I had for the Federal Capital district, a district where 80 per cent of all corporations are registered, I computed typical capital/output ratios relating total physical capital (book values and reassessment) to total sales in each industrial group. Then I multiplied

these coefficients by each firm's sales and obtained in that way an estimate of physical capital. This method assumes that all firms in the group have the same 'capital/output' ratios and that all of them are operating in that year at the same level of plant utilization. This, no doubt, is a somewhat risky assumption, but we should keep in mind that this is used only for constructing stratum boundaries.

3 *Importance of the corporate sector*

In order to estimate the importance of the corporate sector within manufacturing I have related the 1963 total sales of that sector to the total production in the same industry group for the same year. The first set of figures comes from the aggregation of sales of individual corporations, and the second set is taken from the Economic Census (1964 for year 1963)

Since the sales of the corporate sector have been taken from balance sheets, which reflect what happened during fiscal periods ending in 1963, while the Census figures cover the calendar year, and since production is not identical to sales, the data from the two sources are not strictly comparable.

The results are shown in column (8) of Table I As we can see, the ratio of corporate-sector output to total production is not uniform among groups. Those that are markedly incorporated are cement, chemicals beer, rubber industries, and car factories. Sectors notably unincorporated are clothing, wood, and furniture manufacturers. I do not include in the last category oil refineries and production of vehicles other than automobiles, because in these areas a substantial part of production is by state owned enterprises On the whole the private corporate sector represents 58 per cent of the total manufacturing production.

4. *The sample*

Bearing in mind that a sample size of 600 was desired and since some problems in getting complete information for all firms could arise, I decided to start with 630 firms. From a careful examination of sources I estimated the number of sample firms with missing information to be around 5 per cent.

Once the entire list of corporations was completed, I ranked the corporations by crudely estimated physical capital from larger to smaller and then divided the population in twenty strata, each containing an equal amount of estimated capital.

All firms falling in strata 1 to 12 were selected, while an equal number of firms was taken from each of the other eight strata. Given that the first 12 strata contained 205 firms, 425 observations remained to be distributed among the other strata. Then 54 units for the 13th stratum and 53 firms

for each of the 14th to 20th sections were taken at random. For some firms it was not possible to obtain the necessary information and they had to be discarded. The number of firms finally included in the sample was 591

Some words have to be added about the characteristics of the sample In particular the number of observations is not the same for all years 1961–7 Some firms started operating after 1961, some others ceased activity in the period 1964–7, and some mergers took place during 1961–7. No firm was discarded for any of these reasons This of course creates some problems of aggregation which are explained later.

In Table I, a description of the sample together with related series are shown In column (3) of this table the number of firms in the sample is presented, while column (4) shows total sales accounted for by these firms in 1963 Column (8) gives an idea of the importance of the corporate sector. Columns (9) and (10) provide information about the percentage that sales of sample firms represent of total corporate sales and of total group production in 1963. Over all, the sample covers 72 per cent of the corporate manufacturing sector and more than 42 per cent of all manufacturing

Once the firms to be included in the sample were selected, copies of their balance sheets for as many years as possible were collected Then I proceeded to obtain for each firm and for each year information on

(1) Acquisition of fixed assets, classified as
- (i) Land
- (ii) Buildings
- (iii) Building equipment
- (iv) Machinery
- (v) Tools
- (vi) Office equipment
- (vii) Cars and trucks
- (viii) Other means of transportation
- (ix) Repairs.

(2) Sales or retirements of fixed assets classified in the same way Also for each firm but only for the years 1961–7.

(3) Inventories

(4) Profit and loss statement items detailed as follows
- (i) Net sales
- (ii) Cost of sales
- (iii) Compensation to directors
- (iv) Net interest and other financial charges
- (v) Rent on titles, shares and bonds
- (vi) Gains or losses from sales of fixed assets

[illegible] total production for each group

No	Industry group (1)	Industrial groups included (2)	Sample: No of firms (3)	Sample: Sales[a] (1963) (4)	Corporate sector: No of firms (5)	Corporate sector: Sales (6)	Production 1963 census[a] (7)	Importance corp sector (6) (7) = (8) (8)	(4) (6) (9)	(4) (7) (10)
1	Meat and meat preparation	201	15	53,005	84	62,334	76,485	0.815	0.847	0.692
2	Dairy products	202	8	9,137	35	12,810	33,305	0.384	0.713	0.274
3	Mill and baked products	205–206	17	32,953	75	38,813	69,734	0.559	0.849	0.472
4	Sugar refineries	207	16	13,824	19	14,316	30,992	0.462	0.960	0.446
5	Cacao, chocolate, and candies	208	4	2,803	18	4,730	6,440	0.734	0.593	0.435
6	Preserved fruits, vegetables, and fish preparation	203–204–209	12	8,032	146	23,353	62,893	0.317	0.344	0.128
7	Distilled alcoholic beverages	211	6	2,802	32	6,058	8,324	0.727	0.463	0.337
8	Wines	212	20	9,955	74	15,018	21,166	0.709	0.663	0.470
9	Beer	213	5	2,850	14	3,848	2,711	1.419	0.741	1.051
10	Soft drinks	214	9	6,319	52	8,593	13,210	0.650	0.969	0.478
11	Tobacco	220	5	21,828	10	22,537	21,759	1.036	0.969	1.003
12	Textile industries, except woven fabrics	231	69	36,922	388	56,428	106,057	0.532	0.654	0.348
13	Woven fabrics	232–233–239	10	3,067	94	7,259	18,104	0.401	0.422	0.169
14	Clothing	241–242–243	8	2,982	146	9,221	35,860	0.257	0.323	0.083
15	Other textile articles	244–249	7	3,809	33	5,478	11,096	0.424	0.695	0.343
16	Wood and furniture	251–252–259–260	6	769	131	5,745	25,537	0.225	0.134	0.030
17	Pulp, paper, and paperboard	271–272–279	28	13,421	114	17,819	26,592	0.670	0.753	0.505
18	Printing and publishing	280	16	7,577	155	14,821	22,992	0.644	0.511	0.330
19	Leather and fur products	241–242–291 292–293–299	6	1,807	87	6,553	18,114	0.362	0.276	0.100
20	Rubber products	300–309	8	20,292	47	22,073	21,196	1.041	0.919	0.956
21	Basic chemical products	311–312	41	21,074	196	30,028	31,451	0.955	0.702	0.670
22	Paint and varnish products	313	6	3,693	38	5,853	8,656	0.676	0.631	0.427
23	Pharmaceutical products and cleaning preparations	319	38	20,691	284	36,767	49,411	0.744	0.563	0.419
24	Oil refineries	321–329	5	44,017	16	45,362	73,491	0.617	0.970	0.599
25	Glass and glassware	332	5	3,130	40	5,639	7,980	0.706	0.555	0.392
26	Pottery	333	5	2,335	30	3,084	3,059	1.008	0.757	0.763
27	Lime and cement	334	4	8,407	8	8,617	9,393	0.917	0.976	0.895
28	Clay construction materials and non-metallic mineral products	331–339	9	2,337	91	5,631	15,493	0.363	0.415	0.151
29	Basic iron and steel products	341	51	25,552	144	29,778	44,280	0.672	0.858	0.577
30	Nonferrous base metals	342	10	5,850	50	8,030	15,430	0.534	0.729	0.390
31	Metallic articles	350	43	8,671	310	16,435	52,430	0.313	0.528	0.165
32	Non-electrical machinery	360–369	26	12,496	289	24,642	54,691	0.450	0.507	0.228
33	Electrical machinery and appliances	370–379	34	17,357	235	27,746	35,853	0.774	0.626	0.484
34	Automobile industry	383	24	62,693	114	67,795	73,343	0.854	0.924	0.854
35	Other vehicles	381–382–384 385–386–389	5	1,449	71	4,318	44,484	0.097	0.336	0.033
36	Scientific instruments and misc manufactured goods	391–399	9	1,803	159	6,676	20,385	0.327	0.270	0.089
37	Diversified	900	1	1,097	1	1,097			1.000	
	Total		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

(vii) Taxes
(viii) Depreciation charges
(ix) Gains and losses other than operational ones
(x) Net gains or losses for the period (net of taxes).

5 *Estimating physical capital*

Rates of return are estimated for each of the years 1961–7 for all groups here considered. Rate of return here is taken simply as the ratio of net profits to physical capital, both appropriately computed I consider first the problem of estimating physical capital

Physical capital is taken to be plant and equipment plus inventories. I deal first with fixed assets (plant and equipment)

Two important facts prevent us from using straight book values to estimate the stock value of fixed assets First, figures for fixed assets as shown in balance sheets are the result of aggregation of purchases of capital goods in different periods, which, because of the long process of inflation experienced in Argentina, are weak estimators of the real figures Second, the rates of depreciation used by firms also are not the most relevant ones For tax reasons firms tend to adopt methods of depreciation that are faster than the corresponding 'true' ones

The method followed here basically consists of taking yearly annual net purchases of capital goods, correcting them for price changes and depreciation, and accumulating them. Thus in computing physical capital I have to some extent ignored stock book values

5a *Price indexes*

The price indexes used in this work to convert money values of net purchases to real values came mainly from three sources One set of indexes comes from the Central Bank.[1] They are implicit price deflators for investment for the period 1935–67, which comprise series for each category of capital goods, except land. For machinery, however, I used Elias's estimates by industry [2]

Both the Central Bank and the Elias indexes are annual, and should be regarded as averages for the year If capital goods acquisition were made uniformly during the year, and if all firms closed their fiscal year in

[1] For the period 1935–50, somewhat less disaggregated Central Bank figures were used, as published in 'Income and product of Argentine Republic' For the period 1950–67, Central Bank implicit deflators for investment were taken from its 'Origin and composition of GNP' and completed for recent years with information from its monthly bulletin

[2] Victor J Elias, 'Estimates of value added, capital and labor in Argentine manufacturing, 1935–1963' (unpublished Ph.D. dissertation, University of Chicago, 1969) Since the Elias indexes are computed only up to 1963, 1964, or 1965 depending upon the major groups, I extrapolated the series, using the general investment implicit price deflator as the auxiliary variable. For sugar industry I used Cordomi's estimates See Manuel Cordomi, 'A study of the production of sugar in Tucuman, Argentina' (unpublished Ph D. dissertation, University of Chicago, 1969)

December we could have used these indexes without significant dating problems Monthly information about capital goods purchases is certainly not available, so an assumption is called for. A reasonable one is to consider all purchases as made in the middle of the firm's fiscal year.

A way to handle the second problem—differences in closing fiscal year dates— is to have monthly indexes. So I decided to compute monthly indexes starting from the annual ones, doing this by means of related series. I took yearly changes as given by the mentioned indexes, and within each year computed monthly variations as if they moved in the same fashion as the general wholesale price index.

For buildings I used the monthly index published by I.N D E C Several unsuccessful attempts were made to build up a price index for land for industrial uses. I finally decided to use the same price deflator for land as for buildings. My impression is that this undervalues the changes of prices in land in urban and suburban areas but in face of the absolute unavailability of data I had no reasonable alternative.[1]

5b Depreciation

As I have mentioned before, corporations in Argentina do not use coefficients of depreciation that accord with the likely 'true' patterns. This fact has implications for the estimates of the value of physical capital as well as for profit figures To discover the way in which true depreciation occurs for each separate item of each group is not an easy problem to solve To undertake a complementary study for this purpose would require as much work as this project itself. Nevertheless, something had to be done to improve upon the depreciation coefficients used by firms.

A possible way of resolving this problem is to use data for the market value of used capital goods of various ages and relate them to values of comparable new goods. It should be recognized, however, that there is not an active market for used capital goods in Argentina, except automobiles

For tax reasons firms want to use, whenever they can, coefficients of depreciation greater than the corresponding 'true' ones. The incentive provided by accelerated depreciation, namely the postponement of payments of the corporation income tax, becomes even more attractive in an inflationary economy, since the savings of taxes is greater the greater the rate of inflation. Neither the income-tax law nor the regulations that come from it speak clearly about coefficients of depreciation. The only provisions are that they should agree with what technically is generally accepted. It is clear that with such a provision there is much room for discretion by a firm.

[1] Investment in land for industries is not of significant importance Among total existing fixed assets in manufacturing, considering only the corporate sector and taking book values, land amounts to around 10–12 per cent in the period 1961–7.

Durability of equipment is certainly not only a technical question but an economic matter as well However, although economic conditions may differ from country to country, the experience of other countries can be used to a certain extent. In face of the lack of information, certain patterns of depreciation must be assumed, imposing a plausible useful life for each class of assets. Thus I decided to use the old (pre-1962) Bulletin F—Depreciation Rate Tables—Useful Lives—of the Internal Revenue Service of the U S Computations were carried out under two different hypotheses first a linear pattern of depreciation, and second, double declining balance for all fixed assets, assuming in both cases the same total durability.

For those kinds of capital goods common to manufacturing as a whole, such as office equipment, buildings, etc , a single length of life was applied to all industry groups. On the other hand, where machinery is specific to a certain group, the corresponding number of years was applied In cases in which Bulletin F shows a range of possible useful lives, a point in the middle was taken For automobiles I have used my own estimates

5c *The problems of revaluation and sales and retirements*

Though sales and retirements of fixed assets are not of great importance for industrial firms, they deserve some attention because the information provided by balance sheets is incomplete. In particular, the balance sheets do not report current market value of sales. There are three problems here, each of which will be discussed.

First, the reassessments that took place in 1961 and 1967 did not bring major problems when dealing with acquisitions of assets However, they do create problems when sales are being considered When an asset is sold that a firm has previously revalued there is no straightforward way of knowing what the *original* value of the asset was Second, the data for sales of fixed assets are not shown separately from the data on the value of assets retired during the fiscal year. Finally, no direct information is provided for the date of acquisition of assets sold

The first problem was solved as follows For each firm it was determined whether the firm revalued its assets either under the 1960 law or under the 1967 law or both, the exact year or years in which the firm registered the reassessment was also entered Then the value of each kind of asset sold in or after the year in which a reassessment was registered was corrected by special coefficients whose main purpose was to restore the original value of these assets

As for the second problem, since no distinction was possible between sales and retirements of assets, I have assumed that one-half of all entries are made up of sales and the other half is made up of retirements

Ignorance of the year in which the asset was acquired is the third problem concerning sales of fixed assets. For a small number of firms taken at random from those that did not revalue, I computed for each class of assets for the period 1961–7 the average age of the assets when they were sold. The results are shown in Table II (figures for land cannot be obtained, so on the average the date of acquisition was assumed to be two years earlier than that found for buildings).

TABLE II

Number of years elapsed between acquisition date and date of sale for fixed assets—average 1961–7

Buildings	3
Building equipment	3
Machinery	4
Tools	3
Office equipment	4
Cars	3

The final computations were made under two different sets of assumptions. first, that all assets were sold at an age equal to half their useful lives, and second, they were sold at an age equal to the average shown in the above table

5d. The problem of heterogeneity of closing dates

Unfortunately for my purposes, in Argentina firms do not have a common closing date, such as 30 June, for their books in each year. The average closing date is in August or September, but there is considerable dispersion in closing dates. I have computed the estimated rates of return for each sample firm for each of *its* fiscal years *ending* in 1961, 1962, 1963 to 1967. I have also estimated rates of return on a calendar basis by taking moving averages of the fiscal-year figures.

5e. Valuation of inventories

Most of the firms in Argentina use the criterion cost or market value, whichever is lower In an inflationary environment, we can reasonably assume that costs are always lower than market values. I assume that all goods have the same turnover rates and acquisitions are made uniformly during the year. Hence, I could estimate the average age of inventory goods and work with the corresponding price-index number.

6. *Profits*

I have made some corrections on profits as shown on the balance sheets in order to have closer approximation to profits accruing to physical capital. To the profit values net of taxes I have added the amount paid as compensation to directors, interest paid, losses due to fluctuations in

exchange rates, losses due to sales of securities, etc., and subtracted interest received, profits coming from investment in other companies, profits from securities and from events not normal to the businesses as such (i e profits from sales of bonds, from exchange-rate fluctuations, and from sales of assets). At the same time, I have added depreciation as computed by the firms and subtracted depreciation charges according to my own computations

A firm, because of its good financial position, may obtain some reduction in the price of inputs because it pays for them promptly, or it may grant longer terms to its customers and include in the price charged for the products an implicit financial charge. The same reasoning holds in reverse cases, i e. firms in a poor financial position.

In a word, I think firms have profits or losses whose origin is financial and which are not reflected directly in the profit-and-loss statements through a particular especial account, but in an indirect way through differences in the prices of inputs and outputs which conceal some (positive or negative) financial charges.

With this in mind, I have computed for each firm its net financial position by summing the value of cash, credits, and deferred charges and deducting debts To the amount of net financial assets I have applied an arbitrary rate of interest that the company is implicitly earning (or paying) on these financial assets

The financial capital market in Argentina shows a structure of interest rates with a high degree of dispersion, so I have to select a rate somewhat arbitrarily. For each period I took a simple average between the net financial assets at the beginning and at the end of the fiscal period and then applied the rate of inflation for the period as though it were the interest rate The rate of inflation for each period was computed on the basis of a 12-month moving average of the general wholesale price index That is to say, I assumed that if a firm had any positive or negative financial position, its earnings (or losses) on the financial assets were at least equal to the rate of inflation. Of course, when my computed interest charges were included, the firm's charges as registered in the financial accounts were not taken into consideration

Two additional issues merit some consideration The first one is advertising. When advertising is done by a firm, its costs are usually written off in the period when the outlays occur, even though some of the advertising may have long-term effects Thus some part of advertising expenditures consists of investment rather than current outlay and in principle it should be treated this way. Another example of this type is found in the case of some outlays on research and development. The same kind of reasoning is again relevant. Some items of current cost are not strictly current but rather produce some benefits in later periods.

In line with these examples is the case of investment in human capital Some firms do tend to develop among their personnel skills that are not general and are called specific to the firm. That is to say firms are willing to train their employees in some techniques and to give them some knowledge that is useful only within a particular firm The expenses for providing this training appear as part of the current cost of production A recent study by Professor L. G. Telser[1] showed a positive correlation between investment in 'specific' human capital and rates of return implying that measured profits include both return to physical capital and human capital

II. The rate of return estimates

1 *Using industry group data*

On the basis of information aggregated at the industry group level I have estimated rates of return for eight different combinations of assumptions The first four estimates assume that the age of assets sold was on the average equal to half their useful life and the second four assume ages of assets sold as shown in Table II In each set of four, estimates were computed for the two methods of depreciation, linear and double declining balance, and with and without correction for imputed interest on financial assets. Results for all combinations of assumptions are shown only for the sample as a whole (Table III) and detailed results for only one of the methods are presented in Table IV

I notice from the examination of the results first that in general the estimates that assume linear depreciation are very similar to those assuming double declining balance depreciation Second, the results are not significantly different between the two alternative assumptions concerning the average age at which the assets were sold (But here the dispersion among rates of return is somewhat higher under the assumption that assets are typically sold at one-half their useful lives than under the alternative assumption.)

When corrections for imputed interest on net financial assets were made, rate of return estimates tended in all cases to be higher than the corresponding unadjusted estimates. This is a consequence of the fact that the industrial sector as a whole shows in all years a net negative financial position The divergence between adjusted and unadjusted rates of return tended to be higher in periods of business contraction.

It is important to notice that in general, for the sample as a whole, estimated rates of return tend to follow the economic cycle fairly closely In Table V and Fig 1 values of the index of industrial production are

[1] L G Telser, 'Some determinants of the return to manufacturing industries', Center for Mathematical Studies in Business and Economics, University of Chicago (Report 6935)

TABLE III

Rates of return under different assumptions for the sample as a whole, 1961–7

	Age of assets sold		Patterns of depreciation		Adjustment for imputed int on financial assets		Rates of return						
	½-useful life	Table II	linear	Double-declining balance	No	Yes	1961	1962	1963	1964	1965	1966	1967
Method 1	×		×		×		0 139	0 113	0 068	0 107	0 158	0 110	0 100
Method 2	×			×	×		0 140	0 097	0 060	0 088	0 152	0 127	0 091
Method 3	×		×			×	0 143	0 175	0 118	0 144	0 188	0 134	0 125
Method 4	×			×		×	0 133	0 163	0 103	0 128	0 185	0 130	0 119
Method 5		×	×		×		0 118	0 097	0·053	0 087	0 137	0 094	0 085
Method 6		×		×	×		0 114	0 089	0 040	0 079	0 139	0 093	0 082
Method 7		×	×			×	0 121	0 153	0 100	0 122	0 165	0·117	0 109
Method 8		×		×		×	0 118	0 149	0 090	0 117	0 170	0 118	0 109

TABLE IV

Method 5. Rates of return—age of assets sold. according to Table II pattern of depreciation linear. No adjustment for imputed interest on financial assets

	Industrial groups	*1961*	*1962*	*1963*	*1964*	*1965*	*1966*	*1967*	*1961–7 mean*	*1961–7 standard deviation*
1	Meat and meat preparation	−0·0626	0·0841	0·0249	−0·1086	−0·0528	−0·0173	−0·1505	−0·0404	0·0793
2	Dairy products	0·0702	0·0832	0·0870	0·1013	0·1053	0·0776	0·0663	0·0844	0·0148
3	Mill and baked products	0·1210	0·1501	0·1488	0·1289	0·1061	0·1107	0·0908	0·1233	0·0220
4	Sugar refineries	0·0520	0·0667	0·0832	0·1169	0·0665	0·0126	0·0090	0·0581	0·0382
5	Cacao, chocolate, and candies	0·0071	0·0702	0·0681	0·0958	0·0984	0·0880	0·0991	0·0752	0·0327
6	Preserved fruits, vegetables, and fish	0·0379	0·1173	0·1168	0·0441	0·0909	0·0881	0·0586	0·0791	0·0328
7	Distilled alcoholic beverages	0·0928	0·1524	0·1156	0·1428	0·1931	0·2448	0·1360	0·1539	0·0507
8	Wines	0·0960	0·1082	0·0709	0·0637	0·1230	0·1377	0·0900	0·0992	0·0257
9	Beer	0·0335	0·0412	−0·0323	0·0079	0·1355	0·0756	0·0877	0·0499	0·0552
10	Soft drinks	0·2005	0·0798	0·0535	0·0746	0·1096	0·0782	0·0877	0·0977	0·0483
11	Tobacco	0·0516	0·0908	0·1003	0·0911	0·0749	0·0528	−0·0146	0·0638	0·0395
12	Textile industries	0·1024	0·0674	0·0436	0·0665	0·1259	0·0883	0·0409	0·0764	0·0310
13	Woven fabrics	0·0982	0·0809	0·0819	0·0778	0·1311	0·1043	0·1123	0·0981	0·0196
14	Clothes	0·2878	0·0592	0·1468	0·1492	0·2512	0·1972	0·1706	0·1803	0·0750
15	Textile articles	0·1462	0·1565	0·1282	0·2086	0·1328	0·0565	0·0700	0·1284	0·0519
16	Wood and furniture	0·1344	0·0778	0·0237	0·0418	0·0880	0·0874	0·0828	0·0766	0·0357
17	Pulp, paper, and paperboard	0·0709	0·0920	0·1158	0·0490	0·0256	0·0321	0·0356	0·0601	0·0340
18	Printing and publishing	0·1824	0·1688	0·1679	0·1607	0·1487	0·1585	0·1782	0·1665	0·0116
19	Leather and fur products	0·2084	0·1132	0·0739	0·1161	0·1515	0·1675	0·1852	0·1451	0·0467
20	Rubber products	0·2624	0·2514	0·1895	0·1760	0·1205	0·0995	0·1542	0·1791	0·0615
21	Basic chemical products	0·1064	0·0711	0·0510	0·0755	0·1016	0·0876	0·1104	0·0862	0·0217
22	Paint and varnish products	0·1628	0·1587	0·1321	0·1256	0·1250	0·1000	0·1108	0·1307	0·0231
23	Pharmaceutical products	0·0864	0·1190	0·1131	0·1322	0·1385	0·1285	0·1332	0·1216	0·0178
24	Oil refineries	0·2670	0·3009	0·1188	0·0950	0·0673	0·1248	0·1256	0·1671	0·0896
25	Glass and glassware	0·1766	0·1608	0·0388	0·0424	0·1835	0·1509	0·0826	0·1194	0·0631
26	Pottery	0·0409	0·1299	0·0725	0·0957	0·1478	0·1441	0·1373	0·1097	0·0411
27	Lime and cement	0·0691	0·0827	0·0586	0·0512	0·0428	0·0213	0·0794	0·0579	0·0217
28	Clay construction materials	0·1928	0·1738	0·1187	0·0929	0·1479	0·1602	0·1648	0·1502	0·0341
29	Basic iron and steel products	0·0273	0·0195	−0·0035	0·0074	0·0375	0·0265	0·1107	0·0322	0·0372
30	Nonferrous base metals	0·0731	0·0922	0·0079	0·0497	0·1489	0·1266	0·0723	0·0815	0·0470
31	Metallic articles	0·1247	0·0768	0·0469	0·0806	0·1175	0·0920	0·0855	0·0891	0·0261
32	Non-electrical machinery	0·1449	0·1132	0·0455	0·0769	0·1113	0·0451	0·1012	0·0912	0·0372
33	Electrical machinery	0·1289	0·0904	0·0544	0·0749	0·0517	0·1001	0·0816	0·0831	0·0268
34	Automobile industry	0·4232	0·1698	0·0339	0·2847	0·5521	0·2903	0·1346	0·2698	0·1769
35	Other vehicles	0·1154	0·0395	0·0074	−0·0207	−0·0747	−0·0249	0·0317	0·0105	0·0603
36	Scientific instruments	0·2081	0·1582	0·1501	0·1502	0·1842	0·1902	0·1468	0·1697	0·0242
37	Diversified	0·2085	0·0541	0·0106	−0·0161	−0·0372	−0·0117	0·0614	0·0385	0·0833
	Total	0·1152	0·0968	0·0535	0·0878	0·1372	0·0941	0·0852		

shown, together with weighted average rates of return (Table IV, last line) for the sample. Rate of return figures are plotted opposite March of each year, corresponding to the typical middle of the fiscal year (September to August). (The correspondence would be a bit closer if index of production were corrected for its trend.)

TABLE V

Index of industrial production and rates of return 1960–8

Index of industrial production Year and Qtr.			*Mean rate of return Weighted (1)*	*Unweighted (2)*	*Standard deviations (3)*	*Standard dev.—mean, (3)—(2) (4)*
1960	I	98 5				
	II	99 7				
	III	99 3				
	IV	102·5				
1961	I	109 7				
	II	111 1	0 118	0 128	0 092	0 72
	III	111 6				
	IV	106 4				
1962	I	119 6				
	II	112 9	0 097	0·111	0 057	0 51
	III	95 8				
	IV	92 8				
1963	I	96 5				
	II	95 2	0 053	0 077	0 052	0 68
	III	100 8				
	IV	107 4				
1964	I	107 5				
	II	111 7	0 08	0 087	0 069	0 79
	III	120 3				
	IV	120 9				
1965	I	127 8				
	II	129 2	0 137	0 115	0 099	0 86
	III	131 5				
	IV	129 8				
1966	I	114 8				
	II	133 5	0·094	0 110	0 069	0 63
	III	133 7				
	IV	135 5				
1967	I	116 5				
	II	136 1	0 085	0 091	0 061	0 67
	III	134 1				
	IV	130 6				
1968	I	120 2				
	II	139·1				
	III	144·9				
	IV	150 5				

Inspection of the industry rate of return estimates also showed quite large dispersion of these rates across industries within each of the years. Column (3) of Table V shows the (unweighted) standard deviation of the rates of return across industry groups calculated from the rate of return figures by industry in Table IV. Notice that the dispersion among

industries follows much the same cyclical pattern as the mean rate of return given in column (1) of Table V. Column (4) shows the corresponding unweighted coefficients of variation, ranging from about 50 per cent to almost 90 per cent, and showing somewhat the same cyclical pattern as both the mean rate of return and the standard deviation

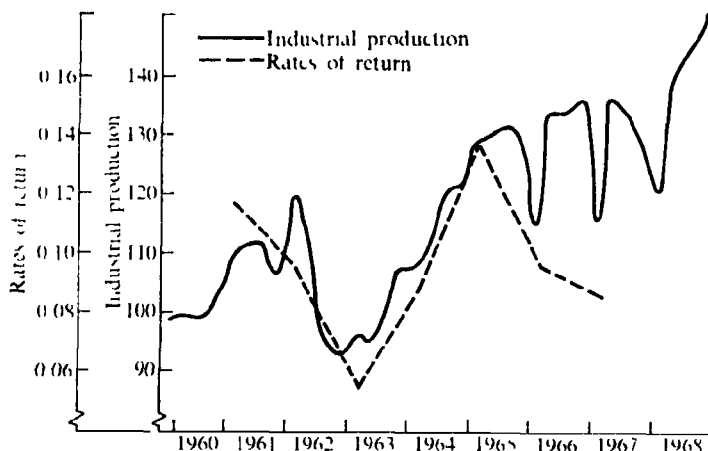



FIG 1 Index of industrial production and rates of return.
SOURCE: Table V

To what extent is it true the industries with high (low) rates of return, compared with the average, in any single year of the period also had high (low) rates of return in the other years? That is, to what extent did the inequality in rates of return persist over the period? Table IV shows for each industry heading the mean of the rate of return estimates over the seven-year period. The inequality among the industries in their seven-year average rates of return is still large. The standard deviation of these seven-year averages is 0 0560 The mean of the seven standard deviations by year in Table V is 0·071. These data indicate that over this seven-year period the tendency toward equality of rates of return was not strong

Evidence of the same kind was obtained when I computed coefficients of correlation across industries of rates of return in pairs of years All the correlations were positive and they averaged about 0·54. (This average went down only to 0 48 when the correlations of rates of return in adjacent years were excluded) Correlation was almost 0·6 between the rates of return for the extreme years 1961 and 1967.

I noted above that there was a strong tendency for the rates of return to vary among years. Table IV shows for each industry heading the standard deviation of the yearly rate of return figures across the seven years 1961–7. The mean of the intertemporal standard deviations over the

thirty-seven industry groups was 0·043, which is not greatly smaller than the corresponding dispersion across industries A substantial part of this intertemporal dispersion no doubt was related to changes in general business conditions in Argentina

2 *Using firm data*

On the basis of information for each sample firm I was able to compute rate of return estimates under the assumption that fixed assets were sold at ages equal to half their useful lives, and for both patterns of depreciation —linear and double declining balance (i e. methods 1 and 2).

When the estimates for firms were averaged at the industry group level, the results, not shown here, were close to those obtained using industry group data. These estimates by industry group were made assuming a common closing month for all firms in the group, taken to be equal to a weighted average of the closing months registered for each firm (with weights equal to sales).

I also estimated from the firm data average rates of return by industry group on a calendar year basis This was done by computing for each industry heading moving averages of the fiscal year estimate described in the preceding paragraph The results do not differ much from those made on the fiscal year basis, but they show a closer correspondence to the business cycle

Since the sample was a stratified random sample with known probability ratios, it is possible to estimate the average rate of return for the population of the corporate manufacturing sector as a whole Thus I computed the average rate of return for each year stratum in the sample, and then a weighted average of these rates across strata with weights equal to the estimated total physical capital in 1963 by strata Table VI shows the estimates by year for the corporate sector. The estimates by strata, not shown here, showed little tendency to differ from one stratum to another.

TABLE VI
Adjusted rates of return for the corporate sector

	Pattern of depreciation	
Year	*Linear*	*Double declining balance*
1961	0 1220	0 1130
1962	0 0885	0 0746
1963	0 0798	0 0638
1964	0 1168	0 1052
1965	0 1419	0 1363
1966	0 1135	0 1051
1967	0 1245	0 1192

3. *Estimating the social rate of return*

To estimate the social rate of return to physical capital in the corporate sector, we start from the private rate and then make the pertinent adjustments so as to take into consideration the divergences between social and private rates of productivity of capital

The most important divergences to be taken into account are taxes, elements of monopoly profits, and imperfections in the labour market. In what follows I have been able to take account only of taxes.

The general idea is to add to private profits those taxes 'paid out' by capital. Direct taxes do not offer serious analytical problems, simply add such taxes to net profits. In the case of indirect taxes, several procedures are available to allocate their amount among the factors of production One approach, that used by Professor Harberger, is to distribute indirect taxes in proportion to the value added by each factor I now describe the methods followed to approximate the social rate of return

3a *Indirect taxes*

The ratio of indirect taxes to value added has been estimated from the Argentine national income accounts. The following table shows how this was done. Column (1) shows income originating in manufacturing (net of indirect taxes) as estimated by the Central Bank. Column (2) shows my corresponding estimates of depreciation charges. Depreciation charges have been computed for each year by taking from the sample data the ratio of estimated depreciation to sales and then applying this ratio to industrial production. Column (3) is the result of subtracting column (2) from column (1) Column (4) is the amount of indirect taxes originating in manufacturing as shown in the national income accounts, and column (5) is the ratio of column (4) to (3), that is the amount of indirect taxes 'paid' by each monetary unit of value added.

3b *Direct taxes*

While the matter of indirect taxes is not an arduous one to deal with, direct taxes pose a more complicated empirical problem On the one hand, national income accounts do not offer any information about direct taxes, and, on the other, there are no complete published series of direct taxes paid by industrial corporations. Two methods of dealing with direct taxes were explored. One was to compute direct tax liabilities on an accrual (legal) basis and the other was based on cash approach for the years for which information on direct taxes paid by industrial corporation could be obtained

The two principal direct taxes existing in the Argentine tax system to which corporations are subject are the income tax and the 'wealth' tax

The first is a proportional tax on profits and the second is computed on the basis of the book value of outstanding shares plus reserves minus investment in other companies. In some years there has been an income-tax surcharge.

To estimate the ratio of profits before taxes to profits after taxes I worked with information published by I.N.D.E.C. for all corporations

TABLE VII

Indirect taxes as a proportion of value added

Year	Income originating in industry[a] (1)	Estimated depreciation charges[a] (2)	Net value added[a] (3)	Indirect taxes[a] (4)	(5) = (4)/(3)
1961	347·2	25·7	321·5	57·5	0·171
1962	417·5	33·9	383·6	62·9	0·164
1963	496·6	48·9	447·7	80·7	0·183
1964	696·3	63·4	632·9	94·6	0·147
1965	1015·7	74·7	951·0	141·6	0·149
1966	1209·7	102·0	1107·7	215·4	0·194
1967	1440·0	127·5	1312·5	326·6	0·248

SOURCE. Central Bank, *Boletín Estadístico* and estimates.
[a] Thousands of millions of pesos m/n.

registered in the Federal Capital district. With the first (legal) method in the case of income tax, I took net profits (after estimating allowances for the deductions for investment credits) and multiplied it by the legal tax rate to estimate the tax liability. In the case of the wealth tax I computed the base as defined by law and then applied the legal rate to the base. This is shown in Table VIII. Columns (1) to (6) are self-explanatory. Net profits as reported by corporations are different from the net profits concept that I use to estimate the rate of return on physical capital since the latter includes all adjustments referred to in Section I. From the information for the firms in the sample I found that the ratio of net profits computed according to my own estimates to the corresponding figure reported by corporations is on the average for the period 1961–7 equal to 1·175. Hence in column (7) the figures of column (6) are adjusted downward by 17·5 per cent.

Alternatively, I estimated the proportion of net profits paid in direct taxes on a cash basis. The procedure was as follows. I started from the data for all corporations registered in the Federal Capital district for which two corrections were necessary. One was to extend the data to the national level and the other to bring it into conformity, with the procedures followed in this work to estimate the rate of return on physical capital. To extend

the value of profits to the national level I assumed that, on the whc ratio of profits to sales was the same for corporations registered (the Federal Capital district as for those registered in the district I simply multiplied each year's profits of Federal district corporati the ratio of total corporate sales to sales of corporations registered Federal Capital district in 1963. The second adjustment is the sa that used to go from column (6) to column (7) of Table VIII.

TABLE VIII

Direct taxes estimated on accrual (legal) basis

(Corporations registered in Federal Capital District)

Year	*Net profits*[a] (*1*)	*Income tax deductions from tax liabilities*[a] (*2*)	*Profits before estimated income tax*[a] (*3*)	*Estimated wealth tax*[a] (*4*)	*Wealth and income taxes*[a] (*5*)	*Ratio of taxes to net profits* (*6*)
1961	28,948	3,729	38,210	1,544	10,806	0.373
1962	23,217	7,200	26,258	1,996	5,037	0.216
1963	21,462	5,013	27,223	2,301	8,062	0.375
1964	42,096	3,985	63,097	2,958	23,959	0.569
1965	69,539	1,140	110,088	4,421	44,970	0.646
1966	75,174		121,150	5,254	51,230	0.681
1967	79,443	6,731	108,525	8,867	37,949	0.477

SOURCE Sociedades Anonimas, *Boletin Mensual*, I N D E C , several issues
[a] Millions of pesos m/n

The amount of income tax paid by industrial corporations was available to me directly by the Internal Revenue Service of Argenti the years 1965 to 1967 The corresponding figure for 1961 was esti on the basis of published information by I N D E C on the amo taxable income for these corporations To estimate on a cash bas amount of wealth tax paid by manufacturing corporations I distr the total wealth tax to the different branches of economic activity (facturing, agriculture, etc.) in proportion to the income tax pa corporations in each of these branches. (Previously I had examin ratio of income to wealth tax basis and had found that it did not significantly between manufacturing and other sectors)

The resulting estimates of direct taxes on a cash basis are gi Table IX

Table X shows my estimates of the social rate of return for manufac corporations The notes to the table explain its derivation from (tables in this section. The social rate of return estimates range 16 per cent to 20 per cent for the cash basis estimates and, for the years from 19 to 25 per cent for the accrual (legal) basis estimate Table XI, I present estimates of social rates of return for several cou

for which such estimates have been made by other persons An effort was made to present the results for periods as close as possible to the one for which I made estimates.

Once Elias's estimates for Argentina are adjusted so as to put them on comparable basis to my own estimates they are very similar Elias used

TABLE IX

Direct taxes on 'cash basis' as a proportion of net profits

Year	Net corporation profits (Fed Cap) (1)	Adjusted industrial corporate profits (2)	Income tax paid by individual corporations (3)	Income tax surcharge (4)	Wealth tax paid by individual corporations (5)	Total direct taxes (6)	Ratio of (6) (2) (7)
1961	28,918	41,685	5,265		1,873	7,138	0·171
1962	23,217	33,432	n a		1,897		
1963	21,462	30,905	n a		2,092		
1964	42,096	60,618	n a		3,040		
1965	69,539	100,136	14,661	2,199	4,169	21,029	0·210
1966	75,174	108,250	8,613	1,291	7,755	17,659	0·163
1967	79,443	114,308	18,732		7,817	26,519	0·232

TABLE X

Estimates of social rates of return

Year	Private rate of return (1)	Rates of return before direct taxes: On accrual (legal) basis (2)	Rates of return before direct taxes: On cash basis (3)	Social rate of return: On accrual basis (4)	Social rate of return: On cash basis (5)
1961	0·1220	0·1607	0·1428	0·1883	0·1673
1962	0·0885	0·0940		0·1090	
1963	0·0798	0·1053		0·1246	
1964	0·1168	0·1733		0·1988	
1965	0·1419	0·2198	0·1717	0·2525	0·1972
1966	0·1135	0·1792	0·1320	0·2039	0·1573
1967	0·1245	0·1750	0·1533	0·2186	0·1931

NOTE (1) From Table VI
(2) Column (1), this table, times [1+column (7) of Table VIII]
(3) Column (1), this table, times [1+column (7) of Table IX]
(4) Column (2), this table, times [1+column (5) of Table VII]
(5) Column (3), this table, times [1+column (5) of Table VII]

fixed assets as a denominator while I used fixed assets plus inventories From sample data I estimated that the value of inventories is on the average equal to one-third of total physical capital. When Elias's estimates are adjusted for the exclusion of inventories, the estimated social rate of return is roughly 0·20

4 *Investment and rates of return*

In this section I present an exploratory examination of the relation between investment and rates of return. The analysis is admittedly somewhat rudimentary.

4a. *Investment and rates of return for the sample as a whole*

In Table XII indexes of gross and net acquisitions of fixed assets for 1961–7 are presented together with my estimates of the private rate of return (methods 1 and 3) and the index of industrial production The investment and rate of return figures refer to the sample as a whole—that is to all of the manufacturing industry groups. These series are graphed in

TABLE XI

Social rates of return estimated in other studies

	Country and author	*Period*	*Rates of return*	*Sectors*
1	Argentina (J. V. Elias)[a]	1955–63	30 9[h]	Manufacturing
2	Brazil (C G Langoni)[b]	1960–7	14 3	Manufacturing
3	Colombia (A C Harberger)[c]	1960–7	11 3–12 4	Private sector less housing
4	Chile (A C Harberger and M Solowsky)[d]	1940–64	15	Whole economy
5	India (A C Harberger)[e]	1955–9	10–19 3 14–26 1[i]	Corporate industrial sector
6	Mexico (M Solowsky)[f]	1940–64	20	Whole economy
7	U S (Stigler and my adjustments)[g]	1951–8	9 61	Corporate manufacturing sector

[a] Victor J Elias, 'Estimates of value added, capital and labor in Argentine manufacturing, 1935–1963' (unpublished Ph D dissertation, University of Chicago, 1969)

[b] Carlos G Langoni, 'A study in economic growth · the Brazilian case' (unpublished Ph D dissertation, University of Chicago, 1970)

[c] Arnold C Harberger, 'On estimating the rate of return to capital in Colombia', Chicago, 1968 (Mimeographed)

[d] Arnold C Harberger and Marcelo Solowsky, 'Key factors in the economic growth of Chile' Paper presented at a Conference on the Next Decade in Latin American Development at Cornell University, April 1966

[e] Arnold C Harberger, 'Investment in men vs investment in machines the case of India', in C A Anderson and M J Bowman (eds), *Education and Economic Development* (Chicago Aldine, 1965)

[f] Marcelo Selowsky, 'Education and economic growth some international comparisons' (unpublished Ph D dissertation, University of Chicago, 1967)

[g] George J Stigler, *Capital and Rates of Return in Manufacturing Industries* (New York Princeton University Press, 1963), and National Income Accounts

[h] Capital includes only fixed assets

[i] Corrected for imperfections in the labour market.

Fig. 2. The figure shows that while rates of return tend to follow the economic cycle, rates of investment do not in general. At first glance the behaviour of total investment during the period is puzzling 1962 was a year of falling activity and relatively low rates of return (unadjusted for implicit financial charges). Nevertheless, investment went up, 1964 and 1965 were years of expansion and relatively high rates of return, but investment stayed low. When we adjust for imputed interest on financial assets the behaviour of investment in 1962 is better explained, but for 1964 and 1965 the picture is still obscure. An interpretation in terms of lags is not of too much help in explaining what happened in these years.

In my opinion a plausible reason for this lies in some important changes that took place during the period in the sphere of tax incentives to investment. Until October 1962 firms could deduct from net taxable profits up to 50 per cent of the value of acquisitions of fixed assets. From this date the deduction was allowed only when the amount of investment was at least 10 per cent of total assets, a condition not very easy to fulfil for most

TABLE XII

Investment in fixed assets, rates of return, and index of production

Year	Index of acquisitions of fixed assets		Rates of return		Index of industrial production
	Gross	Net	Unadjusted	Adjusted for imputed interest	
1961	140	150	0.139	0.143	109.6
1962	159	162	0.113	0.175	98.7
1963	138	129	0.068	0.118	96.3
1964	83	79	0.107	0.143	109.2
1965	81	60	0.158	0.188	127.3
1966	78	60	0.110	0.134	127.9
1967	136	137	0.100	0.125	130.5

of the firms. The announcement of this change could have led to a bunching of investment followed in subsequent years by a marked decline in investment. In 1965 all tax incentives of this kind were terminated, which could have worked to keep investment low in spite of growing industrial activity and higher rates or return, especially if entrepreneurs expected the suspension to be temporary.

In 1967 fiscal incentives were again granted to investment. This time up to 100 per cent of the value of investment was allowed to be deducted from taxable profits, on the conditions that the fixed assets were new and of domestic origin. Notice that investment turned upward sharply in 1967.

Let me point out that the investment behaviour in Fig. 2 also can be observed in general in the industry groups. I computed gross investment indexes for each of the separate groups and they behaved in much the same way as the aggregate index. In addition, I examined the data for the whole corporate sector as published in I.N.D.E.C. In order to have figures not influenced by price variations and number of firms, I took the ratio net purchase of fixed assets to total sales of the same year. The results are shown in Table XIII together with corresponding figures for my sample of corporate firms. The I.N.D.E.C. investment sales ratio is fairly similar to that for the sample and both move in a manner somewhat similar to the investment indexes in Table XII.

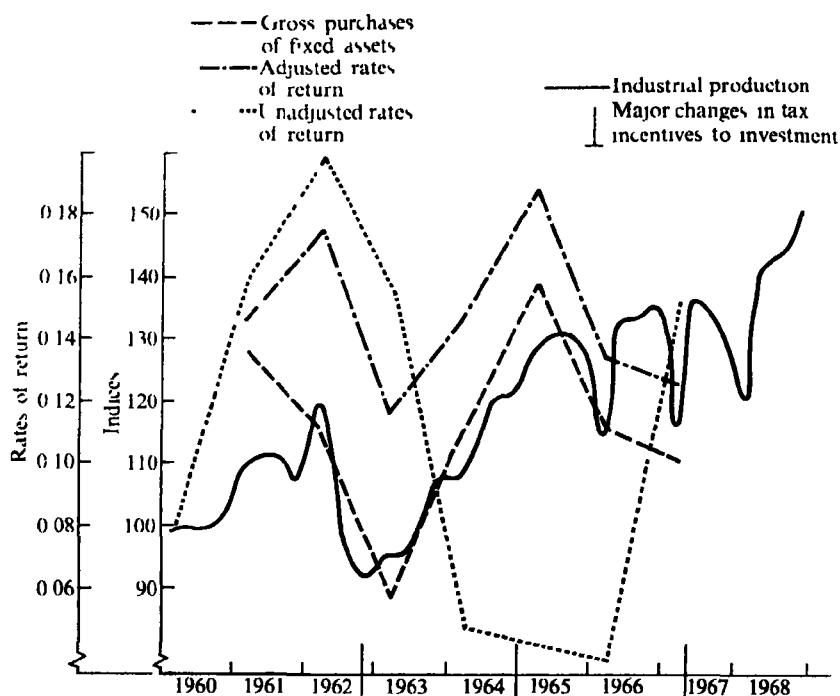



Fig. 2. Investment, rates of return, and index of industrial production.

Table XIII

Investment in fixed assets and sales in the corporate sector

	All corporations—Federal Capital district			*Sample*		
Year	*Sales*[a]	*Investment in fixed assets (excl. buildings)*[a]	*Investment/sales*	*Sales*[a]	*Investment in fixed assets (incl. buildings)*[a]	*Investment/sales*
1960	287,727	15,129	0.0525			
1961	396,735	25,053	0.0631	333,783	44,150	0.132
1962	468,865	42,209	0.0900	405,571	59,580	0.147
1963	550,831	46,402	0.0829	496,806	60,020	0.121
1964	819,722	36,669	0.0447	679,996	41,120	0.061
1965	1,219,441	44,039	0.0361	955,486	48,170	0.054
1966	1,539,717	44,508	0.0298	1,185,702	60,720	0.051
1967	2,000,027	81,573	0.0478	1,494,536	160,990	0.114

Source: I.N.D.E.C., *Boletín de Estadística*, Sociedades Anónimas, several issues.
[a] Millions of pesos m/n.

4b *Some exploratory regressions*

For each of the separate industry groups I computed:

a Total net investment for the period 1961–7 expressed as a fraction of initial physical capital in 1961. (There were two estimates—one based on linear depreciation and the other based on double declining balance depreciation.)

b. Corresponding total gross investment for the period 1961–7

c Corresponding total gross investment in fixed assets for 1961–7

d The average net rate of return over 1961–7 (Again there were two estimates corresponding to the two depreciation assumptions)

e The corresponding gross rate of return figures.

f The coefficient of variation of the net rate of return figures by year over the seven years 1961–7

g The similar coefficient of variation for the yearly gross rate of return figures

These figures used only the information for 498 firms for which data were available for the whole period 1961–7. Furthermore, all of the rate of return figures include the adjustment of profits for imputed financial charges[1] and assume ages of assets sold as shown in Table II.

Table XIV shows the results of regressing investment (net or gross) on the average rate of return (net or gross) across industries All of the regression coefficients are positive and statistically significant The best results obtain in the regressions of gross investment on the gross rate of return

TABLE XIV

Investment and rates of return

(Profits corrected for imputed financial charges)

Dependent variable	*Independent variable*	*Regression coefficient*	*t*	R^2	*F*
1 Total net investment (linear)	Net rate of return on physical capital (linear)	2 396	3 673	0 278	13 497
2 Total net investment (double declining balance)	Net rate of return on physical capital (double declining balance)	2 192	3 595	0 269	12 929
3 Total gross investment (linear)	Gross rate of return on physical capital (double declining balance)	3 476	4 108	0 325	16 884
4 Total gross investment (double declining balance)	Gross rate of return on physical capital (double declining balance)	3 784	4 458	0 362	19 874

Table XV shows similar regressions with an additional explanatory variable, the intertemporal coefficient of variation of the rate of return. The additional explanatory variable, whose regression coefficient is negative as expected, increases the regression coefficients (and their *t*-ratios) for the rate of return variable and increases the coefficient of determination (R^2) Again the results are best for gross investment and the gross rate of return

In Table XVI gross investment in fixed assets is regressed on the *net* rather than gross rate of return. Here too, all of the regression coefficients

[1] I also computed similar rate of return figures without this correction, but they gave much poorer results than the corrected figures in regressions similar to those reported here

TABLE XV

Investment and rates of return

(Profits corrected for imputed financial charges)

Dependent variable	Independent variables		First variable		Second variable			
	1st variable	*2nd variable*	*Regression coefficient*	*t*	*Regression coefficient*	*t*	R^2	*F*
1 Total net investment (linear)	Net rate of return on physical capital	Coefficient of variation of variable 1	2 691	4 101	−0 049	−1 745	0 338	8 665
2 Total net investment (d d b)	Net rate of return on physical capital	Coefficient of variation of variable 1	2 275	3 776	−0 067	−1 461	0 312	7 742
3 Total gross investment (linear)	Gross rate of return on physical capital	Coefficient of variation of variable 1	3 885	4 607	−0 070	−1 910	0 390	10 906
4 Total gross investment (d d b)	Gross rate of return on physical capital	Coefficient of variation of variable 1	3 895	4 646	−0 098	−1 147	0 400	11 366

TABLE XVI

Investment and rates of return

(Profits adjusted for imputed financial charges)

Dependent variable	Independent variables		First variable		Second variable			
	1st variable	*2nd variable*	*Regression coefficient*	*t*	*Regression coefficient*	*t*	R^2	*F*
1 Gross investment in fixed assets (linear)	Net rate of return (linear)		5 456	4 256			0 341	18 121
2 Gross investment in fixed assets (d d b)	Net rate of return (d d b)		5 996	4 158			0 331	17 292
3 Gross investment in fixed assets (linear)	Net rate of return (linear)	Coefficient of variation of variable 1	6 153	4 871	−0 116	−2 143	0 419	12 289
4 Gross investment in fixed assets (d d b)	Net rate of return (d d b)	Coefficient of variation of variable 1	6 178	4 310	−0 144	−1 327	0 363	9 716

for the net rate of return are positive and statistically significant and the values of R^2 are similar to those in Table XV [1]

These exploratory results indicate that there has been some tendency for investment to be allocated to industries where the rate of return is relatively high and that highly fluctuating rates of return tend to deter investment.

[1] Regressions similar to those in Table XVI were run with total gross investment, rather than gross investment in fixed assets, as the dependent variable. The results in general were poor

III. Summary and conclusions

The purpose of this work is to estimate rates of return to physical capital in manufacturing industries in Argentina for each of the years 1961-7 Physical capital was defined to include fixed assets and inventories

The estimates were based on a random sample of approximately 600 corporations taken from a list of 3,831 that included almost all industrial corporations in Argentina On the whole the private corporate sector in Argentina in 1963, the base year, accounted for 58 per cent of total manufacturing production and the sample itself comprised more than 72 per cent of total corporate sales, that is to say that the sample represents more than 42 per cent of all manufacturing sales

The principal source data consisted of corporation balance sheets which in most cases were secured for the entire life of the sample corporations

The value of fixed assets for each of the years of the period 1961 7 was computed on the basis of annual past investment (classified by the main categories of capital goods) corrected for price changes and depreciation Two facts prevented me from using book values reported by corporations to estimate the stock value of fixed assets the long process of inflation experienced in Argentina and elements of accelerated depreciations in the corporation's accounting practices

Over the period as a whole the net private rate of return for the corporate sector averaged about 0 11 The rate of return estimates within industries tended to vary substantially over the period 1961-7, generally in a manner correlated with movements in the general business conditions There was much variation in rates of return among industries in given years and a marked tendency for industries with high rates of return in any one year to have high rates of return in other years of the period

I also made estimates of the social rates of return, basically according to Professor Harberger's concepts The estimates of the social rate of return for the corporate sector in Argentina for the period 1961-7 ranged from 0 17 to 0·20 when taxes were estimated on cash basis These results are comparable in general to social rates of return estimated in other studies on underdeveloped countries

The last part of Section II presents some exploratory work on the relation between investment and the rate of return. The results suggest that there was some tendency in Argentina for investment to move toward the areas where high rates of return prevailed Fluctuations in the rate of return through time seemed to have had a deterrent effect on investment

Universidad Nacional de Córdoba, Argentina

BIBLIOGRAPHY

1. Banco Central de la República Argentina. *Origen del Producto y Composición del Gasto Nacional.* Buenos Aires, June 1966.
2. —— *Boletín Estadístico*, several issues.
3. Banco Central de la Republica Argentina. *Producto e Ingreso de la República Argentina, 1935–1954.* Buenos Aires, 1955.
4. CORDOMI, MANUEL. 'A study of the production of sugar in Tucuman, Argentina.' Unpublished Ph.D. dissertation, University of Chicago, 1969.
5. ELÍAS, VÍCTOR JORGE. 'Estimates of value added, capital and labor in Argentine manufacturing, 1935–1963.' Unpublished Ph.D. dissertation, University of Chicago, 1969.
6. HARBERGER, ARNOLD C 'Investment in men versus investment in machines the case of India.' *Education and Economic Development.* Edited by C A. Anderson and M J Bowan, Chicago Aldine, 1965
7. —— 'On estimating the rate of return to capital in Colombia.' Published as 'La Tasa de Rendimiento de Capital en Colombia', *Revista de Planeación y Desarrollo* (Oct 1969)
8. —— and SELOWSKY, MARCELO. 'Key factors in the economic development of Chile.' Paper presented at the Next Decade of Latin American Economic Development Conference, Cornell University, Apr. 1966
9. I N D E.C. (Instituto Nacional de Estadística y Censos) *Boletín Mensual,* several issues.
10. —— *Censo Nacional Económico*, Buenos Aires, 1968, 1970
11. LANGONI, CARLOS G. 'A study in economic growth. the Brazilian case' Unpublished Ph.D dissertation, University of Chicago, 1970.
12. SELOWSKY, MARCELO 'Education and economic growth: some international comparisons.' Unpublished Ph D. dissertation, University of Chicago, 1967.
13. STIGLER, GEORGE J *Capital and Rates of Return in Manufacturing Industries* New York: Princeton University Press, 1963.
14. TELSER, LESTER G. 'Some determinants of the returns to manufacturing industries.' Center for Mathematical Studies in Business and Economics, University of Chicago, Report 6935.
15. U.S. Department of Commerce, Office of Business Economics 'The national income and product accounts of the United States, 1929–1965. Statistical Tables' Supplement to the *Survey of Current Business*, 1966.

MARKET STRUCTURE AND INDUSTRY PERFORMANCE: THE CASE OF KENYA

By WILLIAM J. HOUSE*

Introduction

PRICE theory has traditionally placed great emphasis on the *structural* features of a market as a guide to the expected performance outcome in that market. Among other indicators, the number of competitors in relation to the size of the market, as measured by an index of concentration, has been singled out to denote the degree of competition or monopoly in the industry. Given some rather restrictive assumptions, there is expected to be a positive relationship between seller concentration and particular indicators of market performance, such as the difference between price and cost or the rate of return on capital. There have been a number of very recent responses to J. S. Bain's call for 'detailed empirical studies which would formulate specific hypotheses on the relation of market structure to market performance and would then test such hypotheses with available evidence'.[1] The respondents related the performance and structure of industries in developed countries and, on the whole, they indeed found a positive relationship between the degree of monopoly and various measures of profitability.[2]

This paper has two goals. First, to the author's knowledge no previous attempts have been made to relate the structural characteristics of industry to its performance in a developing country, where foreign competitors play such a large role in many industrial markets that their presence cannot be ignored as it has in almost all previous studies of monopolistic market structures in developed countries.[3] Hence, a major task has been to incorporate the influence of foreign competition in the concentration index. The results of such a study should be of special interest to legislators seeking a competitive environment, where such policy instruments as trade licensing, investment incentives, import quotas and tariffs could all be

* An earlier draft was presented to the Institute for Development Studies seminar, University of Nairobi.

[1] J. S. Bain, 'Relation of profit rate to industry concentration: American manufacturing, 1936–1940', *Quarterly Journal of Economics*, Aug. 1951, p. 293.

[2] H. Michael Mann, 'Seller concentration, barriers to entry and rates of return in thirty industries, 1950–1960', *Review of Economics and Statistics*, vol. 48, 1966, p. 296; K. D. George, 'Concentration, barriers to entry and rates of return', *Review of Economics and Statistics*, vol. 50, 1968, p. 273; N. R. Collins and L. E. Preston, *Concentration and Price Cost Margins in Manufacturing Industries* (Univ. of California, Berkeley, 1968); S. A. Rhoades, 'Concentration, barriers and rates of return: a note', *Journal of Industrial Economics*, Nov. 1970, p. 82; H. Michael Mann, 'The interaction of barriers and concentration: a reply', *Journal of Industrial Economics*, July 1971.

[3] One recent study which explicitly took account of foreign competition was L. Esposito and F. F. Esposito, 'Foreign competition and domestic industry profitability', *Review of Economics and Statistics*, Nov. 1971.

manipulated to bring about a greater degree of competition, if the existing degrees of monopoly were shown to lead to excessive profit margins

Second, it brings evidence to bear on the controversy raging in the literature over whether the relationship between performance and concentration is continuous and whether concentration alone partly explains performance or whether barriers to entry exert an independent influence on performance in addition to concentration [1]

The study has attempted to estimate the importance of certain structural variables as an explanation of the differences in performance of manufacturing industries in Kenya The analysis is based on the data contained in the 1963 Census of Industrial Production[2] in Kenya.[3]

The hypotheses

The basic hypothesis of the paper is that price-cost margins will be higher the further removed an industry's structure is from the competitive model The difference between price and cost is taken to measure industry performance while industry concentration is used as a measure of the degree of monopoly A number of assumptions are required to generate a testable hypothesis from this proposition

If total costs include normal profit then the difference between price and cost would be zero in competitive industries and would increase, according to demand and cost conditions, as the degree of monopoly increased For the price-cost margins to act as indicators of the degree of monopolistic or competitive industry performance it is essential to take account of capital costs and to assume that average variable costs are constant in each industry Available cost data include current and estimates of the depreciation part of capital costs but do not include opportunity or interest costs In the main, the study utilizes gross price-cost margins which are inclusive of capital costs so that these margins could be expected to be higher in the more capital-intensive industries for this reason alone This proposition was tested [4]

[1] Bain and Mann assigned industries to either of two concentration categories, namely, 'concentrated' and 'unconcentrated' and found a significant difference between the average performance of each category They also concluded that barriers to entry independently influenced performance Collins and Preston were able to find a significantly continuous relationship between performance and structure as did Rhoades in his recent paper, although the latter has called into question the independent influence of barriers to entry on performance In his reply, Mann appears to have successfully rebutted Rhoades's criticism

[2] *Kenya Census of Industrial Production 1963* (Ministry of Economic Planning and Development, Statistics Division 1965)

[3] The analysis is restricted to 1963, the year of the last published industry census It is proposed, at a later time, to undertake a similar exercise using the, as yet unpublished 1967 Census of Industrial Production and to make comparisons with the results obtained for 1963 However, the 1967 Census contains a much greater degree of aggregation of industries, a problem which awaits resolution

[4] Since the reported depreciation charges in the Census were calculated by applying the prescribed rates of allowable deductions for tax purposes they are very unlikely to reflect

Because of differences in the elasticity of demand for final products it might be that two monopolized industries had different price-cost ratios. Hence, for the purpose of testing the hypothesis that, for a given cost structure, an industry with a higher revenue-cost ratio more closely resembles the monopoly performance than an industry with a lower ratio, it is necessary to assume that industries' demand functions do not differ so greatly in price elasticity that any price-cost differences could be attributable to this cause

Industrial structures are usually ranked according to a concentration index, from single-firm monopoly to many-firm competition However, in the context of a developing country, or, indeed, of any open economy, the usual concentration ratio which attributes x per cent of industry sales or employment to the largest three or four domestic firms would be of limited significance when the contribution of imports to total sales is very large The sole domestic producer in an industry would be accorded a concentration ratio of 100 per cent by the usual reckoning, yet this would grossly over-represent any market influence he might have over price if his sales made up only 10 per cent of domestic market sales, the remaining 90 per cent being imports Therefore, it was necessary to incorporate the influence of foreign producers in the concentration ratios which were then related to price-cost margins

In a small, developing country such as Kenya, it could be that some industries export a large percentage of output to very competitive overseas markets where they have little control over price, no matter how concentrated they might be at home For this reason the proposition was tested that price-cost margins are inversely related to the proportion of industry output exported, since the larger an industry's overseas sales the closer its performance might be expected to resemble the competitive one

In addition to concentration, Bain and Mann found certain barriers to entry, such as economies of scale, product differentiation, and capital requirements, to have an independent influence on industry performance [1] Because of the lack of information about such variables in Kenya this study has been constrained to only relating price-cost margins to the capital output ratio which is expected to be positively related to such margins because a high capital requirement can constitute an important deterrent to potential entrants to an industry Thus the capital output

the true cost of using the capital assets Therefore, it was felt preferable, for the most part, to work with gross price-cost margins, rather than deduct the arbitrarily calculated cost of capital to obtain net price-cost margins However, some use is made of net margins In addition, if the rate of return on capital were equal in all industries, then the absolute amount of 'normal' profit will be larger the more capital-intensive the industry

[1] As already mentioned, Rhoades, op cit, has disputed this conclusion, and his criticisms have been rebutted by Mann. This particular controversy continues

ratio plays the dual role of acting as an indicator of the size of capital costs which must be covered by the gross price-cost margins, as well as serving as a guide to the height of barriers to entry into an industry.

The variables to be measured

The study utilizes the 1963 Census of Industrial Production in Kenya which has a 3-digit classification of thirty-eight industries. However, a number of these industries were dropped from the analysis. The Meat Products industry is almost entirely dominated by the Kenya Meat Commission, which is a Statutory Board appointed by the Minister for Agriculture and is a non-profit-making body. For this reason, Meat Products was excluded. Miscellaneous Foods and Miscellaneous Chemicals were excluded because of their heterogeneous nature and lack of reliable data while the Non-Electrical Machinery industry was dropped because most of its data are estimated aggregates with little breakdown due to the large non-response in the industry. The Shipbuilding and Repairing industry and the Railway Rolling Stock industry were excluded, 50 per cent of the former and 100 per cent of the latter industry being owned by the nationalized and non-profit-making East African Railways and Harbours Board. The Textiles industry and the Clothing industry are combined into one industry since 'the products and materials used overlap to such an extent that the two industries are dealt with together'[1] in the main body of the Census report.

Profitability or performance is measured by the difference between average price and average cost, expressed as a percentage of average price. This price-cost margin, which includes the depreciation charge and 'normal profit', is calculated thus.[2]

$$\frac{\text{Price}-\text{Cost}}{\text{Price}} = \frac{\text{Gross Production} - \text{Industrial Costs} - \text{Non Industrial Costs} - \text{Labour Costs}}{\text{Gross Production}}$$

In a much-quoted passage Scitovsky has written. 'Monopoly and oligopoly consist of a power relation among the sellers or the buyers in a certain

[1] *Kenya Census of Industrial Production*, 1963, p. 46

[2] The main source of data was Appendix Table 1 of the *Census of Industrial Production 1963*. In Table 1 some of the industries' value added and net output data were aggregated and these were allocated in what seems to be economically justifiable, in the ratio of their labour costs. The Census defines 'Gross Production' as 'the value of sales plus the net increase in stocks of work in process and finished goods'. Meanwhile 'value of sales' includes 'the value of sales of goods produced and work done. The value is ex-factory or workshop and excludes cost of delivery ... it also excludes excise taxes.' 'Industrial Costs' means 'cost of materials used in production, plus fuel costs, plus the cost of work given out to sub-contract plus repair and maintenance work', while 'Non-Industrial Costs' are defined as 'all current costs except labour costs, industrial costs and depreciation'.

market, and this power relation depends largely on the number and size distribution of the competing sellers or buyers. Measures of concentration try to express the number and size distribution of competitors in terms of a one-parameter index, which could then be regarded as a direct measure of the degree of oligopoly.'[1] Hence, the strength of this 'power relation' of the oligopolists is expressed via the concentration index. How has this index been constructed for Kenya?

The vast majority of the previous studies undertaken have measured concentration by the percentage of industry output in value terms attributable to the top three or four or eight firms in the industry. However, in a developing country, where the number of firms engaged in manufacturing activities is necessarily small, official data sources are loath to reveal information that could be easily attributed to one or two firms. For this reason the first part of the concentration index that has been constructed is necessarily restricted to establishment data.[2] The measure adopted incorporates the percentage employment of each industry attributable to the largest three establishments in the industry.[3]

The precedent for using employment and plant data to measure concentration is found in Rosenbluth's 'Measures of Concentration'[4] where he concludes that 'analysis of the Canadian statistics ... shows that the ranking of industries by firm-concentration index is very similar to the ranking by plant-concentration index. The Spearman correlation coefficient for the two rankings is 0.947. This analysis is based on employment concentration.'[5] In addition, 'output and employment concentration are highly correlated, so that the value of one can be used with great confidence for estimating the other ... while in general, concentration in terms of fixed assets exceeds output concentration, which in turn exceeds employment concentration, the ordering of industries by concentration level is much the same, no matter which standard of size is used, so that the results of cross-section analysis based on one measure will also be applicable to the others'.[6]

If the concentration index is to express the strength of the market power of oligopolists it is important in the Kenya market to incorporate the

[1] Tibor Scitovsky, 'Economic theory and the measurement of concentration', in *Business Concentration and Price Policy* (Princeton University Press, Princeton, 1955), p. 109.

[2] It is as well to remember Gideon Rosenbluth's remark that 'the set of dimensions actually used will depend only partly on what is most appropriate and very largely on the statistics that are available. In every empirical study of concentration the investigator will have to substitute what he can get for what he would like', G. Rosenbluth, 'Measures of concentration', in *Business Concentration and Price Policy*, p. 84.

[3] These data are taken from Appendix Table 16 (e) of the *Census of Industrial Production*, p. 122.

[4] G. Rosenbluth, op. cit., pp. 57-99.

[5] Ibid., p. 85.

[6] Ibid., p. 92. These conclusions are based on U.S., Canadian, and British data.

influence of foreign competition in the index. The 3-establishment concentration index has been multiplied by the percentage of total Kenya market sales (home gross production plus the value of imports) attributable to Kenya domestic production. The implicit assumptions are that the larger is this percentage and the larger is the 3-establishment concentration index the greater is market power, which requires that all domestic production be sold in Kenya.[1] However, explicit account is taken of the proportion of domestic production exported which, if as seems plausible the Kenya producers face competitive markets overseas, would be inversely related to price-cost margins.

Of course, it could be argued that where the domestic producers in industry A have 100 per cent of the domestic market, while in industry B they have 10 per cent, it does not necessarily follow that the entrepreneurs in industry A have more monopoly power than those in B, because of the threat of competition from potential imports. This threat restricts the ability of producers to raise prices untempered. However, it seems reasonable to assume that the domestic producers in an industry where imports are already significant would have less leeway to raise prices, because of the threat of even greater imports, than the producers in an industry where imports are zero, unless the price in this industry has been raised to the margin, where any small rise in price would be import-inducing.[2]

In the case of a homogeneous product such as sugar, where the 3-establishment employment concentration index is 100 per cent yet domestic production is only 37 per cent of total Kenya market sales, the revised hybrid-concentration index is 37 per cent (i.e. 100 per cent $\times$ 37 per cent). Where the 3-digit industry classification incorporates rather heterogeneous sub-industries use is made of the information in the 1963 census on the value of imports at the 4-digit level.[3] For example, in the industry classified as Paper and Products the 3-establishment employment concentration

[1] Essentially, the ratio which is being sought is

$$\frac{\text{Sales of three Largest Plants}}{\substack{\text{Total Sales}\\ \text{(Domestic + Imports)}}} = \frac{\text{Sales of three Largest Plants}}{\text{Total Domestic Sales}} \times \frac{\text{Total Domestic Sales}}{\substack{\text{Total Sales}\\ \text{(Domestic + Imports)}}}$$

where the first ratio on the right-hand side is approximated by employment data.

[2] A similar argument could be made against the more usual concentration ratios which ignore imports in the domestic market. There, the restraining influence on concentrated industries comes from the threat of the new entrants into the industry. The constraints on their ability to do so would be labelled 'barriers to entry', just as the ability of importers to enter the domestic market is reduced by such barriers as trade licences, import quota, tariffs, and transport costs. In addition, international markets are not necessarily competitive. It may be the case that the domestic industry is dominated by a branch of a large multi-national corporation whose policy is not to allow the products of its overseas branches to compete in the domestic market.

[3] *Census of Industrial Production*, pp. 15–100. These data are derived from the Annual Trade Reports.

index was 58 per cent. The value of domestic production was £1.5 million while the total value of imports under this industry-heading was £3 million, yet the author estimates from the 4-digit data in the Census that only £0.55 million of total imports were competitive with Kenya's domestic production. Despite the large value of imports, it is estimated that Kenya producers held 73 per cent of the market for the line of goods they produced, so that the revised concentration index falls by a small amount to 42 per cent (i.e. 58 per cent × 73 per cent). Such an amendment has been made for all thirty-one industries.

It is generally recognized that the amount of money required to set up an efficient plant or firm can deter new entrants to an industry. This capital requirements barrier stands as a proxy for over-all 'barriers to entry', since this was the only variable it was found possible to quantify under this heading. In reporting depreciation for 1961 the Census classified these charges by three types of assets, namely Buildings, which carried a tax-deductible allowance of 4 per cent of the original cost, Transport equipment where the allowance was 25 per cent of the written down value of the asset, and Other Equipment, where the allowance was 12½ per cent of the written down value. For the latter two categories it was a simple matter to estimate the written-down value of the assets but in the case of Buildings such an estimate was made difficult because the ages of the assets were unknown. Since these assets were being written off over twenty-five years, it was assumed that the average age of the assets was twelve and a half years, which allowed an estimate to be made of the current book-value of the assets. The ratio of the total book-value of assets to gross production in 1961 was then used as the measure of the capital : output ratio.[1] The testable hypothesis is that the larger is this ratio in one industry relative to another the larger is the relative capital requirement barrier in that industry.

It could be argued that, since the price-cost margins are observations for one year only, high margins may be the result of short-run changes in demand which, over time, would be eroded by the competitive adjustment process. Or technical progress may have occurred in some industries, resulting in their having high margins that would be eroded in time as the new techniques are adopted and costs fall in other industries. To the

[1] The gross production data for 1961 are found in ibid., Appendix Table 2, p. 103, and Appendix Table 24, p. 152 contains the depreciation charges. The possible errors introduced into the estimates of the capital stock by arbitrarily assuming the average age of buildings to be twelve and a half years would not be significant given the relatively small proportion in total depreciation of depreciation on buildings. In only five out of thirty-one industries did depreciation on buildings constitute more than 20 per cent of total depreciation. When the average age of buildings was assumed to be seven or eight years old, in only three industries did the size of the estimated total capital stock increase by more than 10 per cent.

extent that these factors have operated high price-cost margins will not be the result of market power as hypothesized. Equally, no account can be taken of the so-called 'expense preference'[1] in which a large part of any monopoly profits would be absorbed by inflated managerial salaries or by expenditures undertaken by management for prestige purposes only

Methods of analysis

Previous researchers in the U.S. have tended to follow one of two different hypotheses Those following the 'distinct break' hypothesis,[2] such as Bain and Mann, purport to have found a significant difference between the performance of highly concentrated industries and all other industries, and from within the highly concentrated category to have found a significant difference in performance between industries with high barriers to entry and those with lower entry barriers Others, such as Collins and Preston, and Rhoades, have sought to find a continuous functional relationship between performance and certain structural variables Collins and Preston found a continuous and significant relationship between performance and average concentration, especially amongst the larger, two-digit industry groups

Rhoades, in disputing Bain and Mann's conclusions about the independent role of barriers to entry in determining profitability, claims that 'it is highly unlikely that an industry characterized by high concentration would have low barriers to entry it seems likely that high concentration and low barriers to entry could exist in an industry in the short run but such situations are the exceptions and are of a transitory nature'[3] Using Mann's data and a linear multiple regression equation relating average rates of return to concentration and a dummy variable representing barriers to entry, Rhoades claimed to find that his barriers to entry variable was insignificant in explaining rates of return In his very recent paper Mann[4] has sought to defend his earlier contention of the independent influence of barriers to entry on profit rates Here, it is hoped to bring some evidence to bear on this controversy from the Kenya experience

For the case of Kenya it was felt desirable to apply both the 'distinct-break' hypothesis and the 'continuity' hypothesis 'The theory of oligopoly is not at present so complete, even at the purely formal level, that we may unequivocally describe either the discrete or the continuous hypothesis as the theoretical expectation Both are worthy of analysis '[5]

[1] This term belongs to Oliver E Williamson, *The Economics of Discretionary Behaviour* (Prentice Hall New Jersey, 1964)
[2] Collins and Preston, op cit, p 105
[3] Rhoades, op cit, p 83
[4] Mann, op cit
[5] Collins and Preston, op cit, p 13

Results of the analysis

1 *The 'distinct-break' hypothesis*

Table I shows the average price-cost margins for industries with a hybrid-concentration index of more than 40 per cent and for those industries with an index of less than 40 per cent For both 'gross' and 'net' margins[1] the average margin of the former industries is distinctly larger than that of the latter and the difference between them is significant at the 1 per cent level in each case From the 'net' margins there is strong evidence for the acceptance of the 'distinct-break' hypothesis

An attempt was then made to establish the independent influence of capital requirements barriers on 'net' price-cost margins However, the difference between the average performances of the 'high' and 'medium' capital requirements industries was not significant at the 5 per cent level, and neither was the difference between the average performance of the industries in the 'medium' and 'low' capital requirements categories [2]

When only the most concentrated industries were examined, the difference between the average performances of the 'high' and 'medium' and between the average performances of the 'medium' and 'low' capital requirements industries were not significant at the 5 per cent level

The 'distinct-break' hypothesis fails to establish the independent influence of the capital requirement barrier to entry on price-cost margins [3]

2 *The 'continuity' hypothesis*

The 'continuity' hypothesis suggests a continuous functional relationship between industry performance and the various measures of industry structure The hybrid-concentration ratio generated here for Kenya is expected to be positively related to price-cost margins as is the capital output ratio since 'gross' price-cost margins are inclusive of depreciation and 'normal' profits Price-cost margins are expected to be inversely related to the proportion of industry output which is exported. A linear functional form is assumed and ordinary least squares applied to the data 'Student's "t"' statistics are in parentheses

The first equations estimated were

$$Pg = \underset{(1{\cdot}09)}{3{\cdot}858} + \underset{(2{\cdot}56)}{0{\cdot}221}Cm + \underset{(2{\cdot}91)}{14{\cdot}738}Kr - \underset{(1{\cdot}03)}{0{\cdot}063}X \qquad R^2 = 0{\cdot}482 \qquad (1)$$

$$Pg = \underset{(1{\cdot}21)}{4{\cdot}738} + \underset{(3{\cdot}44)}{0{\cdot}277}Cm \qquad R^2 = 0{\cdot}292 \qquad (2)$$

[1] 'Net' price-cost margins are calculated by including the 1961 depreciation charges, which were the only ones reported in the 1963 Census, as part of costs However, these 'net' margins still are inclusive of 'normal' profits

[2] Industries were classified as having 'high' capital requirements if the estimated ratio of book value of assets to gross production for 1961 was 0 45 or above, as 'medium' if the ratio fell between 0 45 and 0 15, and as 'low' if the ratio was less than 0 15 As a result the thirty-one industries were almost equally divided between these three categories

[3] Again, it should be emphasized that the 'net' margins still include 'normal' profits

TABLE I

Gross price-cost margins and net price-cost margins for thirty-one manufa turing industries of Kenya, classified into those with a hybrid-concentratu ratio above 40 per cent and those with a hybrid-concentration ratio of le than 40 per cent

Industry	*Price-cost margin (%)*	
	Gross	*Net*
Above 40%		
Canned fruit and vegetables	1 4	0 5
Grain mill products	15 5	12 7
Spirits	50 8	43 9
Beer and malt	32 7	26 6
Soft drinks	26 3	21 2
Tobacco	17 1	10 7
Cordage, rope, and twine	16 5	12 1
Footwear	24 0	19 4
Other wood products	21 2	12 0
Paper and products	26 4	24 7
Tanning and leather	13 7	12 3
Rubber products	12 2	7 4
Basic industrial chemicals	31 5	28 9
Paints	6 2	5 3
Soap	14 3	12 5
Glass and products	32 9	21 6
Cement	30 3	17 4
	22 1	17 0
Below 40%		
Dairy products	7 2	5 0
Bakery products	5 4	3 5
Sugar	12 6	5 4
Confectionery	7 9	6 6
Textiles and clothing	11 0	9 2
Sawn timber	12 9	6 2
Furniture and fixtures	14 6	10 7
Printing and publishing	9 9	7 2
Clay and concrete	16 9	6 3
Metal products	14 8	12 9
Electrical machinery	16 2	14 7
Motor vehicles	10 5	9 7
Motor repairs	8 1	7 2
Miscellaneous manufacturing	11 1	5 8
Average	11 4	7 9

SOURCE *Kenya Census of Industrial Production* 1963

where Pg = 'Gross' price-cost margin 1963

Cm = Hybrid concentration index, account being taken of imports

Kr = Ratio of book value of assets gross production 1961, as a measure of the ratio of capital output

X = The proportion of an industry's output exported.

The F-test applied to equation (1) shows the three variables to be jointly significant at the 1 per cent level and together they explain 48·2 per cent of the variation in gross price-cost margins, the coefficient of Cm being significant at the 2 per cent level and the coefficient of Kr being significant at the 1 per cent level. Although the sign of the coefficient of X is negative, as expected, it is not significant.[1]

Equation (2) shows that Cm alone is able to explain 29·2 per cent of the variation in gross margins. The coefficient is significant at the 1 per cent level.

Equations (3) and (4) fulfil our expectations that in an economy such as Kenya's where imports of manufactured goods are large, plant concentration explains little of the variation in gross price-cost margins.

$$Pg = 7{\cdot}167 + 0{\cdot}060Cp + 18{\cdot}394Kr - 0{\cdot}018X \qquad R^2 = 0{\cdot}387 \qquad (3)$$
$$\qquad (1{\cdot}91) \quad (0{\cdot}99) \qquad (3{\cdot}43) \qquad (0{\cdot}28)$$

$$Pg = 9{\cdot}923 + 0{\cdot}118Cp \qquad R^2 = 0{\cdot}116 \qquad (4)$$
$$\qquad (2{\cdot}42) \quad (1{\cdot}92)$$

where Cp = the simple index of plant concentration, i.e. percentage of employment in three largest plants.

The explanatory variables in equation (3) are jointly significant at the 1 per cent level while only Kr is significantly different from zero, and this at the 1 per cent level. In equation (4) R^2 is very low and the coefficient of Cp is significant only at the 10 per cent level.

As explained earlier, where an industry exports a large proportion of its production to competitive markets overseas, a high concentration index at home would not signify market power, as hypothesized, and would not lead to high price-cost margins.[2] Therefore, it was felt that a better representation of the importance of these structural variables on price-cost margins might be obtained by excluding seven industries which exported

[1] Multicollinearity would appear to present no serious problems since the simple correlation coefficient between Cm and Kr was 0·402, between Cm and X it was 0·435 and between X and Kr it was 0·065.

[2] For example, the Canned Fruit and Vegetables industry was one of the most concentrated by the author's calculation, since Cm stood at 76 per cent, yet this industry exported 90 per cent of its production and its gross price-cost margin was 1·4 per cent, the lowest of the thirty-one industries.

more than 50 per cent of their production overseas The exercise was then repeated for the remaining twenty-four industries and the results were

$$Pg = 0{\cdot}959 + 0{\cdot}384Cm + 7{\cdot}699Kr - 0{\cdot}099X \qquad R^2 = 0{\cdot}621 \qquad (5)$$
$$(0{\cdot}27) \quad (4{\cdot}02) \qquad (1{\cdot}50) \qquad (1{\cdot}09)$$

$$Pg = 3{\cdot}03 + 0{\cdot}415Cm \qquad R^2 = 0{\cdot}563 \qquad (6)$$
$$(0{\cdot}09) \quad (5{\cdot}23)$$

The F-test applied to equation (5) showed the variables to be jointly significant at the 1 per cent level and the equation suggests that Cm, Kr, and X are able to explain 62 per cent of the variation in Pg while they explained only 48 per cent when the export oriented industries were included In equation (5) only Cm is significant (at the 1 per cent level) yet the simple correlation coefficient between Cm and Kr for the twenty-four industries is 0·507, suggesting multicollinearity could be a problem and that in equations (5) and (6) the independent influence of Cm on Pg is exaggerated and the importance of Kr is underestimated

The next step was to examine the evidence on whether capital requirements barrier to entry exerts an important influence on performance independent of the concentration index. The above exercise was repeated for all thirty-one industries using the 'net' price-cost margin as the dependent variable If the coefficient of the Kr variable were positive and significantly different from zero then there would be some evidence for the independent influence of the capital requirements barrier on performance The estimated equations were

$$Pn = 3{\cdot}238 + 0{\cdot}254Cm + 3{\cdot}824Kr - 0{\cdot}081X \qquad R^2 = 0{\cdot}282 \qquad (7)$$
$$(0{\cdot}82) \quad (2{\cdot}61) \qquad (0{\cdot}67) \qquad (1{\cdot}18)$$

$$Pn = 2{\cdot}889 + 0{\cdot}232Cm \qquad R^2 = 0{\cdot}230 \qquad (8)$$
$$(0{\cdot}75) \quad (2{\cdot}91)$$

The variables in equation (7) were jointly significant at the 1 per cent level, and together they explain 28·2 per cent of the variation in 'net' price-cost margins but only Cm is significantly different from zero, and it is significant at the 2 per cent level Alone, in equation (8) Cm explains 23 per cent of the variation in 'net' price-cost margins [1]

On this evidence, the capital requirements barrier to entry appears an unimportant influence on 'net' price-cost margins and the significance of Kr in equation (1) can be attributed to its role as the determinant of that part of price-cost margins which represents depreciation costs and 'normal profit.

[1] As already noted, the correlation coefficient between Cm and Kr for the original thirty one industries is 0 402, not high enough to suggest multicollinearity is too serious a problem

Conclusions

The test of the 'distinct-break' hypothesis shows that there is a positive relationship between gross price-cost margins and monopoly power, as measured by the 'hybrid' concentration index, as well as gross margins and the ratio of book-value of assets to gross production. These results are confirmed by the test of the 'continuity' hypothesis

The proportion of industry production exported was found to be inversely related to price-cost margins as expected, but the coefficient was not significant When the seven industries that exported more than 50 per cent of their production were dropped from the analysis the explanatory powers of concentration, the capital output ratio, and the proportion of industry production exported for the variation in gross price-cost margins increased markedly.

The test of the 'distinct-break' hypothesis failed to reveal any independent influence of the capital requirements barrier to entry on price-cost margins This result was confirmed by the test of the 'continuity' hypothesis where the coefficient of the capital·output ratio is not significant when the net price-cost margin is the variable to be explained [1] However, this variable is significant in explaining gross price-cost margins since they include depreciation charges and 'normal' profits

The results presented here suggest further areas of research effort When the results of the 1967 Census of Industrial Production become available it is proposed to undertake a similar exercise, in order to establish whether such market power and the high price-cost margins persisted in the intervening years and whether the industries identified as holding market power in 1963 maintained such power through these years Furthermore, a closer examination of government licensing, quota and tariff policies in respect of the concentrated industries is required as well as an assessment of the presence and policies of multi-national corporations in these industries It could well be that, if the authorities in Kenya are concerned about the creation of a competitive industrial environment some form of monopolies review body might be instituted to investigate and appraise the policies and practices of the most concentrated industries. Alternatively, more imports might be encouraged to compete with the products of these industries by liberalization of the machinery of import restriction [2]

University of Nairobi

[1] No doubt there are other 'barriers' to entering the Kenya market as exemplified by the attempt of Rothmans of Pall Mall (Kenya) Limited to enter the Tobacco and Cigarettes industry in 1966 The company was formed with a capital of £600,000 with the intention of challenging the British American Tobacco Company (Kenya) Limited, the sole domestic producer During 1967, a fierce fight for the East African market took place, but within the year Rothmans were defeated and sold their assets to B A T for half the original price

[2] Of course, these claims overlook the claims of 'infant' industries

APPENDIX AND INDUSTRY NOTES

The data in Table II were generated from the Census of Industrial Production (Kenya, 1963.

TABLE II

'Gross' and 'net' price-cost margins (Pg *and* Pn), *plant-employment concen tration ratios* (Cp), *gross home production as a percentage of total hom production plus imports* (Hp), *hybrid-concentration ratios* (Cm), *proportio of home production exported* (X), *and book value of assets: gross productio ratios for 1961* (Kr)

Industry	*Pg* %	*Pn* %	*Cp*	*Hp*	*Cm*	*X*	*K*
Dairy products	7 2	5 0	29	92	27	43	0 1
Canned fruit and vegetables	1 4	0 5	96	79	76	90	0 0
Grain mill products	15 5	12 7	43	96	41	14	0 2
Bakery products	5 4	3 5	38	96	36	11	0 1
Sugar	12 6	5 4	100	37	37	1	0 5
Confectionery	7 9	6 6	100	28	28	18	0 0
Spirits	50 8	43 9	100	82	82	11	0 5
Beer and malt	32 7	26 6	60	97	58	18	0 5
Soft drinks	26 3	21 2	41	99	41	5	0 3
Tobacco	17 1	10 7	100	75	75	53	0 [illegible]
Cordage, rope, and twine	16 5	12 1	96	65	62	40	0 4
Textiles and clothing	11 0	9 2	52	35	19	77	0 1
Footwear	24 0	19 4	100	76	76	72	0 4
Sawn timber	12 9	6 2	10	98	10	19	0 [illegible]
Other wood products	21 2	12 0	85	61	52	46	0 8
Furniture and fixtures	14 6	10 7	18	96	17	26	0 [illegible]
Paper and products	26 4	24·7	58	73	42	49	0 1
Printing and publishing	9 9	7 2	23	81	19	6	0 [illegible]
Tanning and leather	13 7	12 3	62	85	53	54	0 1
Rubber products	12·2	7 4	61	73	45	66	0 [illegible]
Basic industrial chemicals	34 5	28 9	60	85	51	88	0 7
Paints	6 2	5 3	86	60	52	37	0 0
Soap	14 3	12 5	66	83	54	46	0 1
Clay and concrete	16 9	6 3	42	88	37	7	0 8
Glass and products	32 9	21 6	100	64	64	34	1 [illegible]
Cement, etc.	30 3	17 4	78	100	78	49	1 0
Metal products	14 8	12 9	44	81	36	49	0 1
Electrical machinery	16 2	14 7	33	100	33	5	0 1
Motor vehicles	10 5	9 7	30	100	30	1	0 0
Motor repairs	8 1	7 2	10	100	10	1	0 0
Miscellaneous manufacturing	11·1	5 8	42	73	31	20	0 4

Industry notes

'Gross' price-cost margins were calculated from Appendix Table I of the *Censu of Industrial Production 1963* and this, together with the information on depreciatio contained in Appendix Table 24, allowed the calculation of 'net' price-cost margi to be made.

The percentage of employment in the three largest establishments was calculate from the information contained in Appendix Table 16 (e), while the information o

mports and exports was gathered from the 'Industry Notes' of the Census, pp 15–100. Data for gross production for 1961 are contained in Appendix Table 2.

For some industries explicit assumptions were required for various reasons and these are outlined below:

Spirits

No data were available for the industry for 1961 since production began in 1963. Therefore, the ratio of book value of assets.gross production for 1961 was used for the beer industry and an estimate made of depreciation for 1963 which was used to calculate the net price-cost margin. Most domestic production was of gin, so only the value of gin imports was considered to be competitive with domestic production.

Tobacco

'Net Output' for the Soft Drinks and Tobacco industries were combined and this figure included £585,000 as the margin from other activities. Since the Tobacco industry distributed its own products, it was assumed these receipts originated from the distribution network. However, the value of production is an estimated value at the factory door so the margin from distribution was deducted from the net output figure, and the remainder was allocated to Soft Drinks and Tobacco in the ratio of their labour costs.

The value of Tobacco imports and exports was not available, but Text Table 41 gave these data in actual quantities, so the proportion exported and the ratio of domestic production to domestic production plus imports were calculated from this Table 41.

Electrical machinery

No depreciation data were given for this industry, so the ratio of book value of assets gross production for 1961 for the Non-Electrical Machinery industry was adopted and an estimate made of depreciation for 1963, which was then used to calculate the 'net' price-cost margin.

THE CLIMACTERIC IN U.S. ECONOMIC GROWTH

By BARRY W. POULSON *and* J. MALCOLM DOWLING

I. Introduction

THE existence of a climacteric or retardation in the trend of economic growth in Great Britain in the latter half of the nineteenth century is supported by a substantial body of literature.[1] Such a consensus is not found with respect to the existence or absence of a retardation in the trend of economic growth in the U S When S J Handfield-Jones and B Weber examined the trend of economic growth in the U.S., they did not find evidence of a climacteric comparable to that experienced in Great Britain [2] They did find evidence of lower rates of economic growth in the first few decades of the twentieth century but interpreted these as a reflection of long swings rather than as a retardation in the trend of growth A number of economists have found evidence of a slowing down in the rate of economic growth in the twentieth century without attempting to distinguish trends from cyclical fluctuation John Kendrick, for example, maintained that the rate of growth of net National Product was subject to progressive retardation from 1890 through the prosperous 1920s.[3] Ed Ames fitted different polynomials to a time series for manufacturing output and identified a break in the trend of growth in the first third of the twentieth century based on a change in the order of the polynomial that best fit the data [4]

Perhaps the most careful attempt to distinguish trends from fluctuations in U S. economic growth is Kuznets's study of capital formation [5] Kuznets used both reference cycle relatives and moving averages to estimate the trend and cyclical components of Gross National Product over the period 1869–1955. He concluded that:

the decadal rates of change in reference cycle averages and in 5 year moving averages

[1] D J. Coppock, 'The climacteric of the 1870's', *The Manchester School of Economic and Social Studies*, vol xxiv, no 1 (Jan 1956), E H Phelps Brown and S J Handfield-Jones, 'The climacteric of the 1890s', *Oxford Economic Papers*, New Series, vol iv, no 3 (Oct 1952), A E Musson, 'British industrial growth during the great depression 1873–76 some comments', *Economic History Review*, Apr 1963, Phyllis Deane and W A Cole, *British Economic Growth 1688–1959* (1962)

[2] S J Handfield-Jones and B Weber, 'Variations in the rate of economic growth in the U.S A, 1869–1939', *Oxford Economic Papers*, vol vi, no 1, Feb. 1954

[3] John Kendrick, *Productivity Trends in the United States*, Princeton University Press for the National Bureau of Economic Research, 1961.

[4] Ed Ames, 'Trends, cycles, and stagnation in U S manufacturing since 1860', *Oxford Economic Papers*, New Series, no 11 (Oct 1959), pp 270–81

[5] Simon Kuznets, *Capital in the American Economy, Its Formation and Financing*, Princeton University Press, for the National Bureau of Economic Research, 1961

describe long swings around a steadily declining rate of growth constituting the long-term trend—although this retardation in the rate of growth cannot be found after the 1920's.[1]

However, Kuznets is ambiguous with respect to the existence of retardation in the trend of economic growth; at another point he states

If we assume that the depression of the 1930's is cancelled by the expansions that preceded and followed it, and regard the long term averages as truly representative secular levels, we find retardation in the rate of growth not only of total income, but also of product per head and per worker. If, however, we omit the depression and the war years and regard the 1946-1955 averages as secular levels, there is no clear case for retardation in the rate of growth of product per capita or per worker.[2]

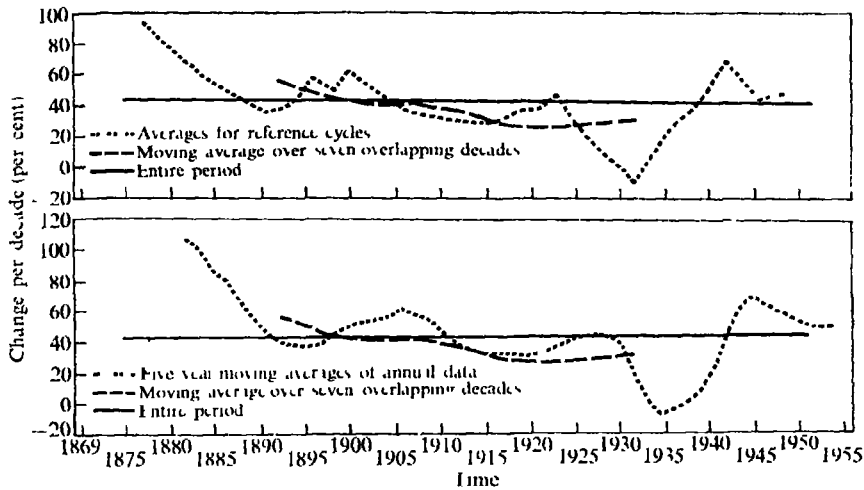



FIG 1 Percentage change per decade in averages for reference cycles and in 5-year moving averages of annual data, Gross National Product, 1929 prices, 1869–1955
SOURCE Simon Kuznets, *Capital Formation in the American Economy*, p. 53

The techniques that have traditionally been used to estimate the trend and to distinguish trend from cyclical components of output have been criticized on a number of grounds If we assume that the trend is a polynomial, as Ames did, we have no *a priori* knowledge of the degree of the polynomial. When the order of the polynomial fitted is insufficient then a portion of the trend will remain in the residuals and be misinterpreted as part of a cyclical fluctuation about the trend When a higher order polynomial is fitted which overadjusts then a portion of the cyclical fluctuation will be contained in the trend estimate. The techniques used by Kuznets to estimate trend have been criticized in a number of studies. Adelman points out that a method relying on reference cycle relatives is very complex and there is no possible analytical evaluation since the *a priori*

[1] Ibid, p 54 [2] Ibid, pp 77–8.

formulation cannot be provided[1] Slutsky has shown that the process of summing and differencing implicit in the moving average technique can introduce spurious cycles in a series which originally contained no cycles[2] Bird *et al* found that this latter procedure also has a tendency to lengthen the mean distance between peaks and troughs of cyclical fluctuations in the series[3] Thus the moving average technique may introduce spurious fluctuations in the data or cause the average duration of cyclical fluctuations in the data to increase making it difficult to separate the cyclical from the trend component

II. Harmonic trend estimation

Granger and Hatanaka have suggested a method of trend estimation which defines trend in terms of frequency and enables us to differentiate the frequencies regarded as trend from other frequencies in the time series[4] The technique they suggest is called harmonic regression analysis In it trend is defined as a sum of periodic sine and cosine functions From a time series containing T observations

$$y_t = a_0 + \sum_{J=1}^{K} (a_J \cos \pi Jt/T + b_J \sin \pi Jt/T) + U_t \quad (t = 1 \quad T), \tag{1}$$

where $a_0 + \sum_{J=1}^{K} (a_J \cos \pi Jt/T + b_J \sin \pi Jt/T)$ is the trend and U_t is the residual The sine and cosines in (1) are the independent variables and the coefficients a_0, a_J, and b_J are fitted by least squares If, for example, $K = 2$ the least squares trend would be given by.

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 \frac{\cos \pi t}{T} + \hat{b}_1 \frac{\sin \pi t}{T} + \hat{a}_2 \frac{\cos 2\pi t}{T} + \hat{b}_2 \frac{\sin 2\pi t}{T}. \tag{2}$$

If $K = 3$ the least squares trend would be given by:

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 \frac{\cos \pi t}{T} + \hat{b}_1 \frac{\sin \pi t}{T} + \hat{a}_2 \frac{\cos 2\pi t}{T} + \hat{b}_2 \frac{\sin 2\pi t}{T} + \\ + \hat{a}_3 \frac{\cos 3\pi t}{T} + \hat{b}_3 \frac{\sin 3\pi t}{T}. \tag{3}$$

Essentially the formulation fits one-half of a complete sine wave to the time series when $J = 1$, a complete sine wave when $J = 2$, one and one-half sine waves when $J = 3$, and so on The amplitude and phases of these

[1] Irma Adelman, 'Long cycles—fact or artifact?', *American Economic Review*, vol 55 pp 444–63

[2] E Slutsky, 'The summation of random causes as the source of cyclic processes' *Econometrica*, vol 5, p 105

[3] Bird, R D, Desai, M J, Engler, J J, and Taubman, P M, 'Kuznets cycles in growth rates the meaning', *International Economic Review*, vol 6, pp. 229–39.

[4] C W. J Granger and M Hatanaka, *Spectral Analysis of Economic Time Series*, Princeton University Press Princeton, 1964

waves are determined by the regression coefficients a_J and b_J Since each set of sine and cosine terms for a given value of J can be interpreted in terms of parts of a sine wave we can think of each sine and cosine term as accounting for the variation in Y_t due to a specific periodic component If for example we have a series of 100 yearly observations then the first harmonic ($J = 1$) accounts for periodic components of 200 years in length (only half the sine curve is completed in 100 years) The second harmonic ($J = 2$) accounts for periodic components of 100 years, the third harmonic ($J = 3$) for components of $66\frac{2}{3}$ years, and so on

The technique suggested by Granger and Hatanaka enables us to define trend in terms of periodic variations with a duration longer than cyclical or long-swing fluctuations in a time series Although the literature is rather ambiguous regarding the duration of long-swing fluctuations most authors define long swings to include periodic fluctuations up to 40 years in duration Therefore we define trend to preclude periodic variations with a duration less than 40 years, i e we adjust the value of K to preclude periodic components of less than 40 years The shortest time series examined in this study has eighty-four observations, therefore the largest value of K considered is 4 which will include a periodic component $\frac{84}{2} = 42$ years For $K = 4$ the least squares trend is given by.

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 \frac{\cos \pi t}{T} \hat{b}_1 \frac{\sin \pi t}{T} + \hat{a}_2 \frac{\cos 2\pi t}{T} + \hat{b}_2 \frac{\sin 2\pi t}{T} + \\ + \hat{a}_3 \frac{\cos 3\pi t}{T} + \hat{b}_3 \frac{\sin 3\pi t}{T} + \hat{a}_4 \frac{\cos 4\pi t}{T} + \hat{b}_4 \frac{\sin 4\pi t}{T}. \quad (4)$$

It is possible that the additional variables in equations (3) and (4) do not add to the explanatory value of the model F tests are used to measure the statistical significance of additional variables in the model It is well known that $F = \frac{\Delta R^2/N}{1-R^2/T-K}$ is distributed as F with N and $T-K$ degrees of freedom ΔR^2 is the incremental change in R^2 when additional explanatory variables are added, R^2 is the coefficient of determination for the model after the variables have been added, N is the number of variables added, T is the number of sample observations, and K is the number of variables (independent plus dependent) in the model after N variables have been added [1] The F value measures the significance of the addition of variables $\frac{\cos 3\pi t}{T}$ and $\frac{\sin 3\pi t}{T}$ when we move from equation (2) to equation (3), and the significance of $\frac{\cos 4\pi t}{T}$ and $\frac{\sin 4\pi t}{T}$ when we move from equation (3) to equation (4).

[1] For example, see D S. Huang, *Regression and Econometric Methods*, Wiley, 1971

TABLE I

Regression coefficients and their t values, equations (2), (3), and (4)

Series	Equation	R^2	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4
GNP	2	0.989303	−113.88	−161.42	−40.13	39.51				
			65.63	17.01	9.72	22.79				
	3	0.992180	−141.90	−87.06	0.50	72.43	17.95	−12.11		
			29.36	2.07	0.02	12.91	6.07	1.71		
	4	0.992788	−90.60	−344.36	−161.16	5.62	−27.73	56.69	15.83	16.10
			4.84	1.60	1.18	0.23	1.69	0.97	0.16	2.81
Non-ag output	2	0.972403	−83.37	−117.76	−32.70	27.94				
			35.21	9.11	5.81	11.82				
	3	0.978607	−114.01	−34.24	12.79	63.91	19.56	−13.39		
			16.66	0.58	0.39	8.05	4.68	1.34		
	4	0.981451	−55.45	329.83	372.45	12.88	−33.61	170.70	−37.96	19.20
			2.14	1.79	1.99	0.38	1.48	−2.13	2.02	2.43
GNP *per capita*	2	0.962324	−1.17	−1.26	−0.38	0.40				
			36.38	7.19	4.97	12.44				
	3	0.964719	−1.36	−0.70	−0.07	0.62	0.12	−0.09		
			13.26	0.79	0.14	5.26	1.97	0.61		
	4	0.971254	0.37	−2.91	−1.43	−1.63	−1.42	0.46	0.11	0.55
			0.99	0.68	0.53	3.37	4.34	0.39	0.42	4.79
Non-farm capital stock (Gallman–Kuznets)	2	0.994408	−197.51	−161.29	−34.92	20.81				
			76.77	11.47	5.71	8.10				
	3	0.998698	−262.87	−330.32	−130.58	97.94	42.51	30.77		
			64.99	9.42	6.74	20.86	17.17	5.19		
	4	0.998794	−271.69	172.82	188.10	109.18	49.88	−106.88	32.53	2.38
			17.32	0.96	1.65	5.37	3.62	2.18	2.83	0.49
Non-farm capital stock (Kendrick)	2	0.984597	−142.51	−66.78	−27.21	7.65				
			39.24	−3.38	3.16	2.12				
	3	0.994034	−186.45	−567.98	−307.40	60.53	29.84	87.83		
			25.22	8.88	8.72	7.06	6.61	8.15		
	4	0.996675	−354.41	−357.59	180.16	279.50	179.86	37.62	−9.92	−53.08
			15.79	1.39	1.11	9.62	9.14	0.54	0.61	7.76
Total labour force	2	0.999200	−31.26	−18.13	−4.97	5.27				
			274.55	22.77	14.33	36.20				
	3	0.999680	−34.30	0.51	5.32	8.83	1.93	−3.13		
			114.07	0.19	3.69	25.30	10.47	7.11		
	4	0.999686	−34.64	19.18	17.15	9.26	2.21	−8.24	−1.21	−0.09
			28.83	1.39	1.96	5.95	2.09	2.20	1.37	0.25
Non-ag labour force	2	0.997257	−28.37	−15.45	−3.59	5.25				
			−108.07	10.82	5.78	20.12				
	3	0.997882	−31.93	−14.69	−3.28	9.43	2.29	−0.02		
			42.47	2.26	0.91	10.81	4.98	0.02		
	4	0.997884	−32.62	−14.65	−3.27	10.33	2.91	−0.007	0.011	−0.22
			10.71	−0.42	−0.15	2.62	1.09	0.0008	0.0005	0.23

Note. All [illegible] coefficients have been scaled down by a factor of 10[illegible]. t values are shown below the regression coefficients.

TABLE II. *Partial F values for models* (3) *and* (4)

Series	*Addition of terms* $\frac{\cos 3\pi t}{T}$ *and* $\frac{\sin 3\pi t}{T}$ (*Changes Model 2 to Model 3*)	*Addition of terms* $\frac{\cos 4\pi t}{T}$ *and* $\frac{\sin 4\pi t}{T}$ (*Changes Model 3 to Model 4*)
GNP	24 0*	3 7*
Non-farm output	11 0*	5 8*
GNP *per capita*	3 5*	11 0*
Non-farm capital stock (Gallman–Kuznets)	16·5*	0 3
Non-farm capital stock (Kendrick)	23 9*	30 0*
Total labour force	80 0*	1 0
Non farm labour force	15 0*	0 001

* Denotes statistical significance at the 90% level

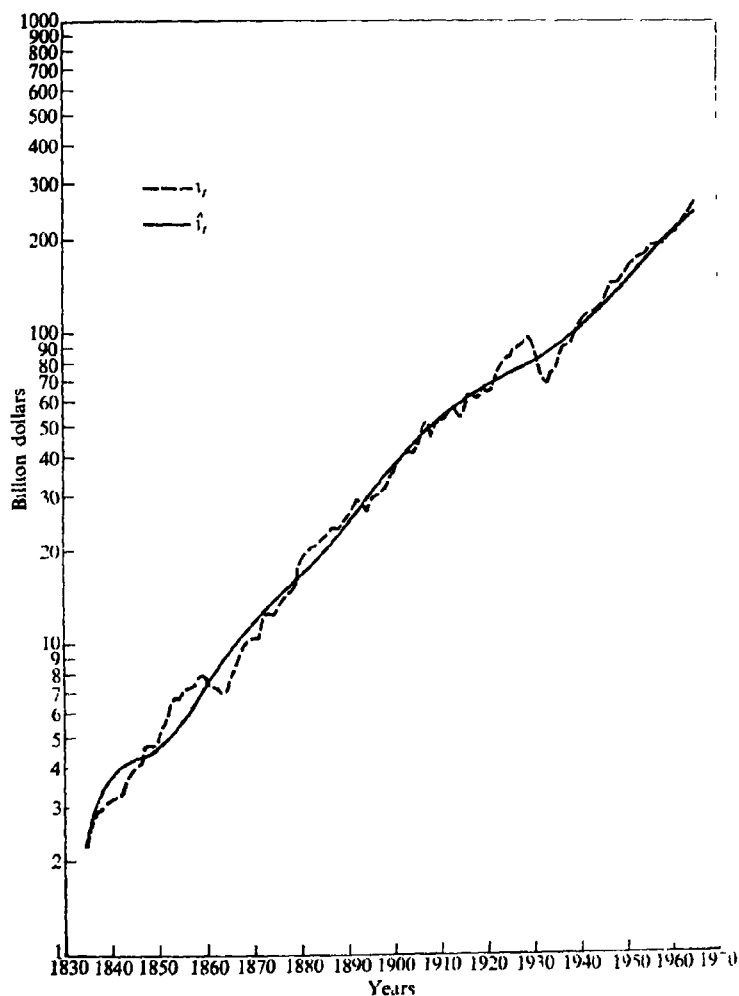



FIG. 2. Harmonic trend for gross national product 1834–1965, 1929 prices

III. Trends in U.S. economic growth

Trends are estimated for total output, total output *per capita*, and no agricultural output to reveal changes in the trend of final output Tren estimated for non-agricultural capital stock and for total labour force ar

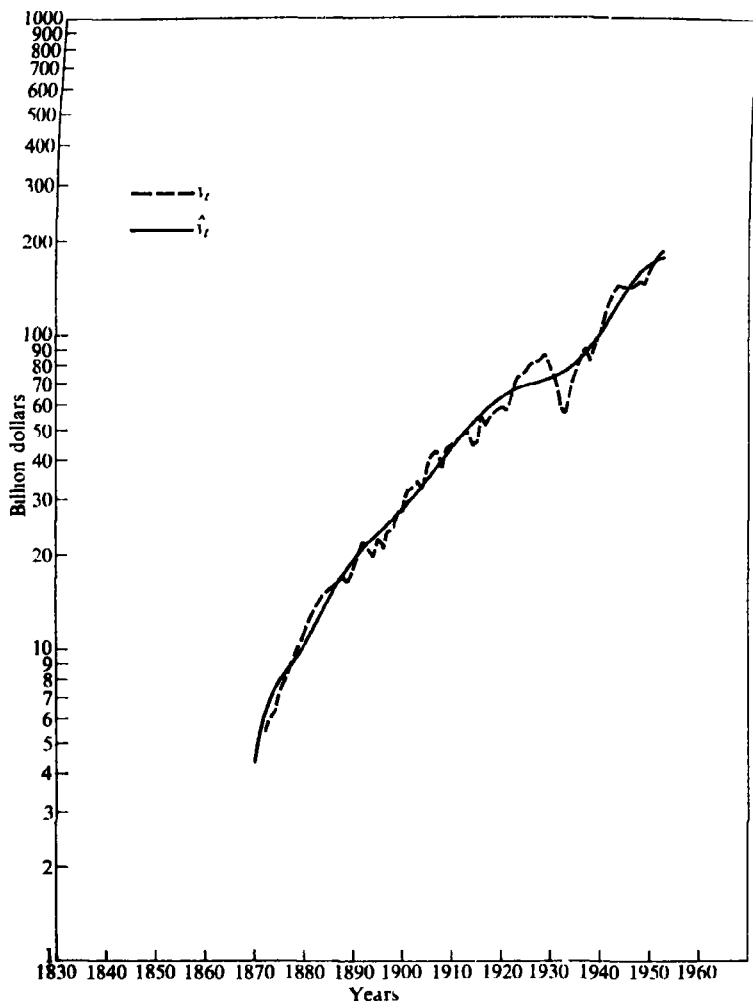



FIG 3 Harmonic trend for gross domestic non-farm product 1869–1953, 1929 prices

non-agricultural labour force, provide evidence of changes in the trend o factor inputs The data sources are listed in the appendix.

Table I lists the R^2, the regression coefficients, and t values for equation (2), (3), and (4). The trend regression equations all explain almost all of the variation in y. This is not unexpected since for most upward trending serie

the low frequency or trend component accounts for most of the total variation in the series.

Table II lists the partial F values for equations (3) and (4) These tests show that in four of the series, output, output *per capita*, non-farm output,

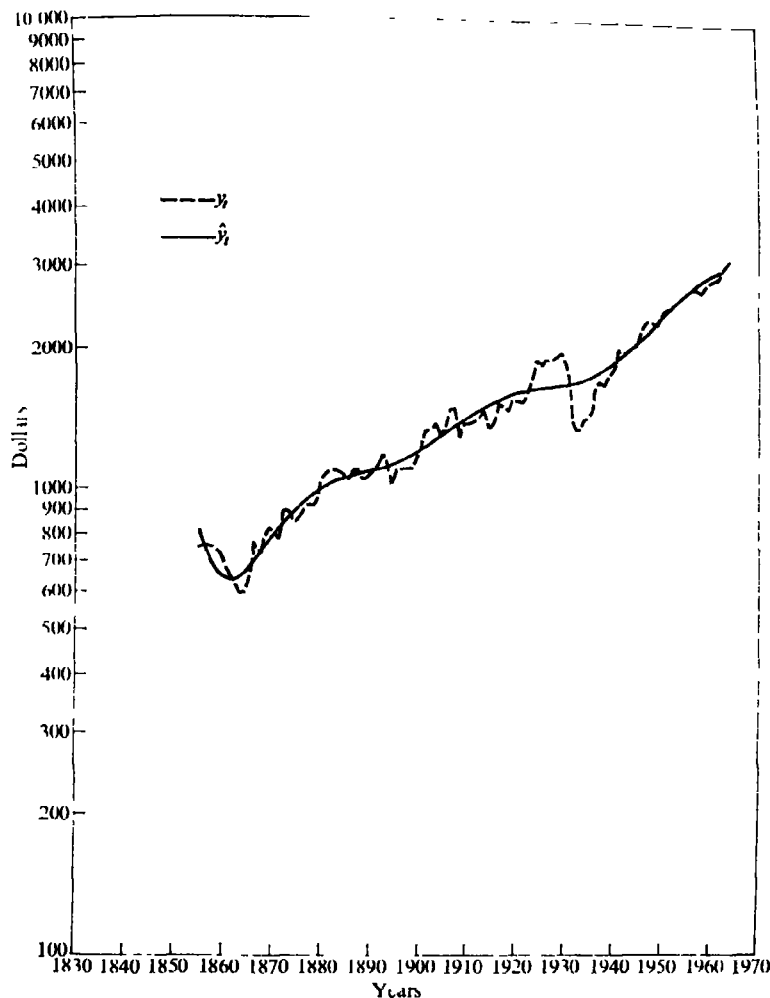



FIG 4 Harmonic trend for gross national product *per capita* 1855–1965, 1929 prices

and non-farm capital (Kendrick's estimates), equation (4) provides the best fit In the other three series, total labour force, non-farm labour force, and non-farm capital stock (Gallman–Kuznets estimates) equation (3) provides the best fit.

Graphs of the logarithms of y_t and $\hat{y}_t$ are shown for the optimally fitted

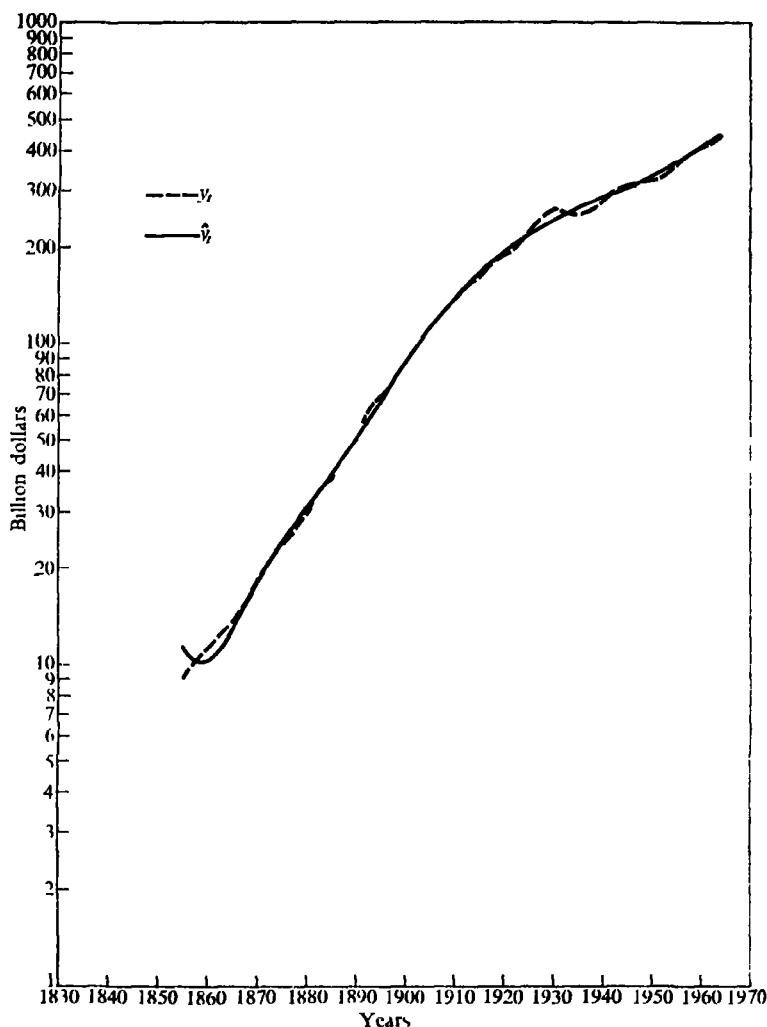



FIG 5 Harmonic trend for gross domestic non-farm capital stock (Gallman–Kuznet 1855–1964, 1929 prices

equations in Figs. 2–7. Changes in the slope of these curves represe changes in the rate of growth in the trend of the series. Casual observati of these curves does reveal a slowing down in the trend rate of change roughly the first third of the twentieth century, the period identified l Kuznets and others as the climacteric in U.S. economic growth. To exami changes in the trend of economic growth more closely we have calculat the rates of change in $\hat{y}_t$ according to the simple rate of change formula

$$\frac{d\hat{y}_t}{dt} = \frac{\hat{y}_t - \hat{y}_{t-1}}{\hat{y}_{t-1}} 100.$$

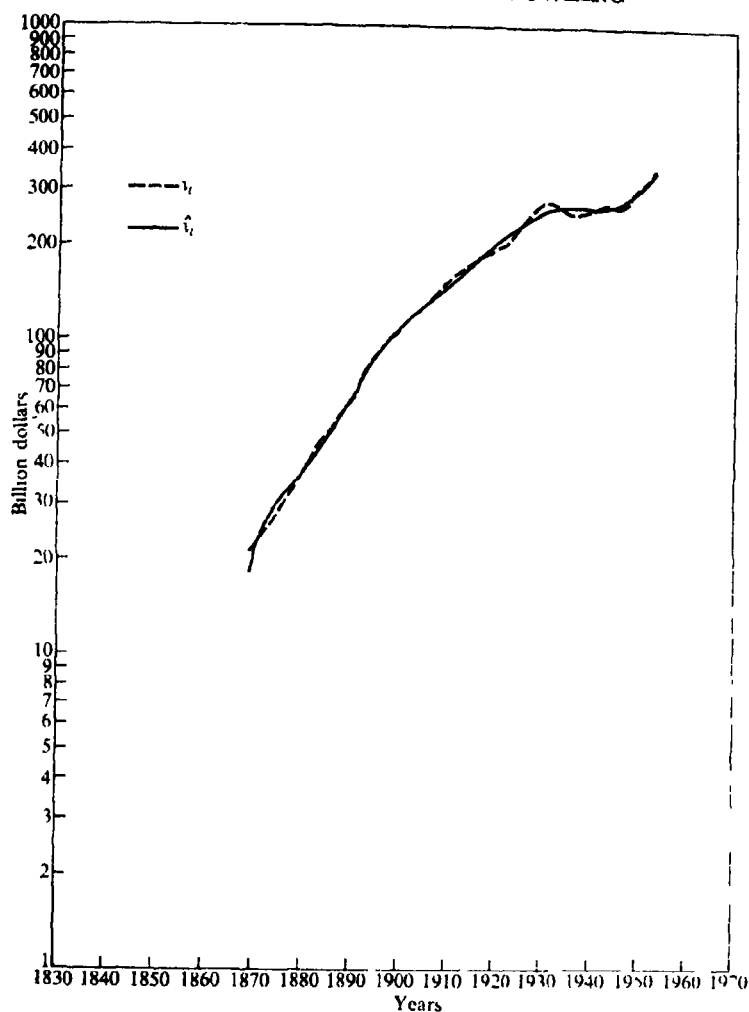



FIG. 6 Harmonic trend for gross domestic non-farm capital stock (Kendrick), 1869-1953, 1929 prices

The annual rate of change in the harmonic trend of each series is shown in Appendix II. These rates of change more clearly identify the retardation in the trend of economic growth. The retardation in trend for output begins in the late 1880s while that for non-farm output and output *per capita* begins in the 1910s. All three output series reach a low point in the rate of change in trend in the late 1920s

Retardation in the trend on non-farm capital stock extends back into the nineteenth century and covers a longer period of time The low point in the rate of change in non-farm capital is not reached until 1939–40 Similarly the retardation in the trend of labour force extends back into the nineteenth

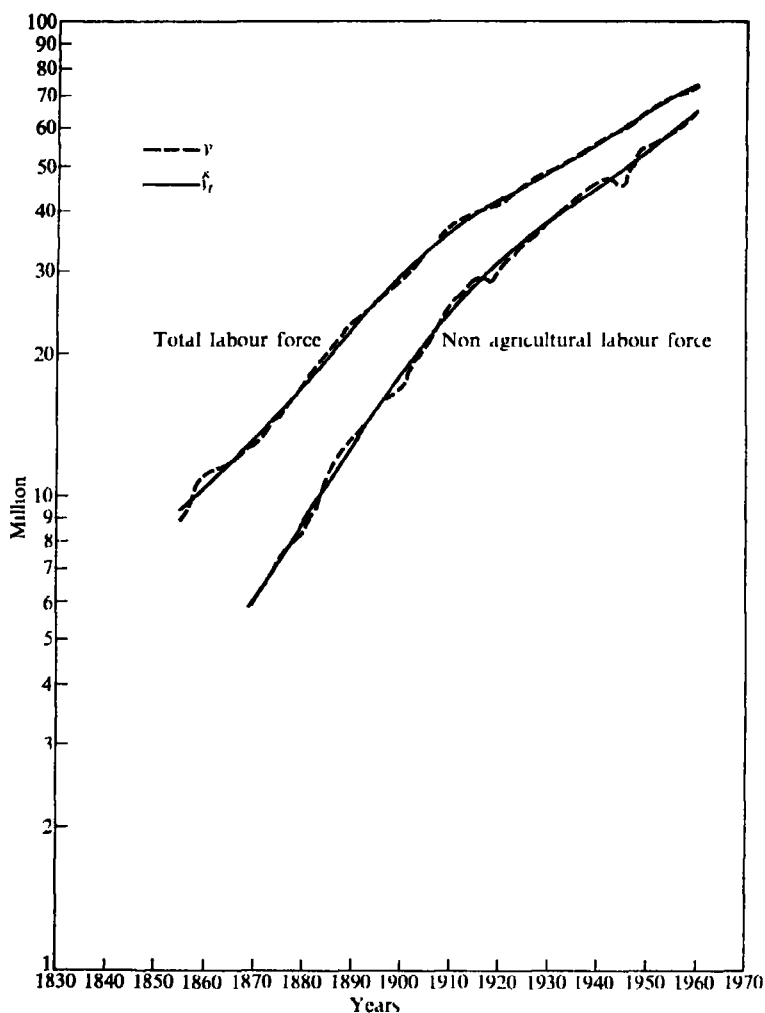



FIG. 7 Harmonic trend for total labour force 1855–1960, and non-agricultural labo force 1869–1960

century. The low point in the trend for total labour force is reached in tl late 20s while the low for non-farm labour force occurs in the late 30s

IV. Conclusion

The literature on economic growth in the U.S. does not provide a bes for categorically accepting or rejecting the thesis that there was a retard tion in the trend of growth. The problem in previous research has been tl difficulty in isolating the trend from business-cycle and long-swing fluctu tions about the trend This study uses harmonic regression analysis to defir

trend in terms of frequency and to distinguish frequencies associated with trend from the business-cycle and long-swing frequencies in a time series

The harmonic trends estimated for all of the time series show evidence of retardation in the first third of the twentieth century The rates of change in the harmonic trends for every series are lower in this period than in any other period during the whole century from 1860 to 1960 The retardation in the trends for output cover the first three decades of the twentieth century. The retardation in the trends for factor inputs cover a longer period of time which extends back into the nineteenth century and generally continue through the first four decades of the twentieth century.

The retardation in the trend of each series reflects periodic variations in the time series with a duration longer than business-cycle and long-swing fluctuations. Therefore, results of harmonic regression analysis lead us to conclude that the U S. did experience a climacteric or retardation in the trend of economic growth in the first third of the twentieth century We also feel that this technique of analysis may be usefully extended to other countries which have experienced economic growth over a comparable period of time.

University of Colorado

APPENDIX I

Sources.

Gross National Product, annual data underlying

Robert E. Gallman, 'Gross National Product in the United States 1834–1909', in *Output Employment and Productivity in the United States after 1800* (New York · NBER, 1966); Thomas S. Berry, *Estimated Annual Variations in Gross National Product 1789–1909*, University of Richmond, Richmond, Simon Kuznets, *Capital in the American Economy*, Princeton University Press, for the National Bureau of Economic Research, 1961 Department of Commerce, *Survey of Current Business*.

Gross National Product *per capita* ·

The above series for Gross National Product divided by a series for population constructed in Ansley J. Cole and Melvin Zelnick, *New Estimates of Fertility and Population in the United States*, Princeton University Press, 1963, pp 21–3

Non-farm output.

John W. Kendrick, *Productivity Trends in the United States*, National Bureau of Economic Research, New York, and Charles Franks, *A Study of Technological Change in the United States non-Farm Economy from 1869–1960*, unpublished Ph.D dissertation, University of Colorado, 1970

Total Labour Force estimates by Stanley Lebergott

Manpower in Economic Growth, McGraw Hill, 1964, interpolated on the population series above for the nineteenth century.

Non-Farm Labour Force estimates

Kendrick, op. cit.

Gross Domestic Non-Farm Capital Stock 1855–1964:

Constructed from the annual series underlying the Gross National Product estimates cited in Fig. 1.

Gross domestic non-farm capital stock 1869–1953:

Kendrick, op. cit.

APPENDIX II

Harmonic Trends

(annual rate of change, dy_t/dt)

Year	*Gross National Product per capita*	*Non-farm product*	*Gross National Product*	*Non-farm capital stock (Gallman)*	*Non-farm capital stock (Kendrick)*	*Total labour force*	*Non-farm labour force*
1835			40 4				
1836			22 9				
1837			14 6				
1838			9 7				
1839			6 7				
1840			4 6				
1841			3 2				
1842			2 2				
1843			1·6				
1844			1 3				
1845			1 3				
1846			1 4				
1847			1 6				
1848			1 9				
1849			2 4				
1850			2 8				
1851			3 3				
1852			3 7				
1853			4 1				
1854			4 5				
1855			4 8				
1856	−6 8		5 0	−6 5		2 7	
1857	−5 8		5 1	−4 4		2 6	
1858	−4 6		5 2	−2 3		2 5	
1859	−3 4		5 2	−1 5		2 5	
1860	−2 1		5 2	1 8		2 4	
1861	−0 9		5 2	3 4		2 3	
1862	0 2		5 1	4 6		2 3	
1863	1 1		5 0	5 6		2 3	
1864	1 9		4 8	6 2		2 3	
1865	2 5		4 7	6 6		2 3	
1866	3 0		4 5	6 8		2 3	
1867	3 3		4 4	6 8		2 3	
1868	3 5		4 2	6 7		2 3	
1869	3 5		4 1	6 6		2 3	
1870	3 5	57 6	4 0	6 4	16 4	2 4	3 0
1871	3 4	26 6	3 9	6 2	12 0	2 4	3 3
1872	3 2	15 2	3 8	6 0	9 1	2 5	3 6
1873	3 0	9 6	3 7	5 8	7 1	2 5	3 8
1874	2 8	6 7	3 7	5 7	5 8	2 6	3 9
1875	2 6	5 3	3 6	5 5	4 9	2 6	4 0
1876	2 3	4 8	3 6	5 4	4·3	2 6	4 1
1877	2 1	4 9	3 6	5 3	4 0	2 7	4 1
1878	1 9	5 3	3 6	5 2	4 0	2 7	4 1
1879	1 6	5 8	3 6	5 1	4 1	2 8	4 1
1880	1 4	6 3	3 7	5 1	4 4	2 8	4 1
1881	1 3	6 8	3 7	5 0	4 7	2 8	4 1

APPENDIX II (*cont.*)

Year	*Gross National Product per capita*	*Non-farm product*	*Gross National Product*	*Non-farm capital stock (Gallman)*	*Non-farm capital stock (Kendrick)*	*Total labour force*	*Non-farm labour force*
1882	1 1	7 0	3 8	5 0	5 1	2 9	4 0
1883	0 9	7 2	3 8	5 1	5 5	2 9	4 0
1884	0 8	7 1	3 9	5 1	5 8	2 9	3 9
1885	0 7	6 9	4 0	5 1	6 1	2 9	3 9
1886	0 7	6 7	4 0	5 2	6 3	2 9	3 8
1887	0 6	6 3	4 1	5 2	6 5	2 9	3 8
1888	0 6	5 9	4 2	5 3	6 5	2 9	3 7
1889	0 6	5 5	4 2	5 3	6 5	2 9	3 7
1890	0 6	5 1	4 3	5 3	6 4	2 8	3 6
1891	0 7	4 7	4 3	5 4	6 2	2 8	3 6
1892	0 7	4 3	4 3	5 4	6 0	2 8	3 6
1893	0 8	4 0	4 3	5 4	5 8	2 7	3 5
1894	0 9	3·8	4 3	5 4	5 5	2 7	3 5
1895	1·0	3 6	4 3	5 4	5 2	2 7	3 5
1896	1 1	3 5	4 3	5 4	5 0	2 6	3 4
1897	1 2	3 5	4 2	5 4	4 7	2 6	3 4
1898	1 3	3 5	4 2	5 3	4 4	2 5	3 3
1899	1 4	3 6	4 1	5 3	4 1	2 4	3 3
1900	1 5	3 7	4 1	5 2	3 9	2 4	3 3
1901	1 6	3 9	4 0	5 2	3 7	2 3	3 2
1902	1 6	4 1	3 9	5 1	3 5	2 3	3 2
1903	1 7	4 3	3 8	5 0	3 3	2 2	3 1
1904	1 7	4 5	3 7	4 9	3 2	2 1	3 1
1905	1 7	4 6	3 6	4 8	3 1	2 0	3 0
1906	1 8	4 8	3 5	4 7	3 0	2 0	3 0
1907	1 7	4 9	3 4	4 6	3 0	1 9	2 9
1908	1 7	4 9	3 2	4 5	2 9	1 9	2 9
1909	1 7	4 9	3 1	4 3	2 9	1 8	2 8
1910	1 6	4 9	3 0	4 2	3 0	1 8	2 8
1911	1 5	4 8	2 9	4 1	3 0	1 7	2 7
1912	1 5	4 6	2 8	3 9	3 0	1 7	2 7
1913	1 4	4 4	2 7	3 8	3 1	1 6	2 6
1914	1 3	4 2	2 6	3 7	3 1	1 6	2 6
1915	1 2	3 9	2 5	3 5	3 1	1 5	2 5
1916	1 1	3 6	2 4	3 4	3 1	1 5	2 4
1917	1 0	3 3	2 3	3 3	3 2	1 5	2 4
1918	0 9	3 0	2 2	3 1	3 1	1 4	2 3
1919	0 8	2·7	2 1	3 0	3 1	1 4	2 3
1920	0 7	2 4	2 1	2 9	3 0	1 4	2 2
1921	0 6	2 1	2 0	2 8	3 0	1 4	2 2
1922	0 5	1 8	2 0	2 6	2 9	1 4	2 1
1923	0 4	1 6	1 9	2 5	2 7	1 4	2 1
1924	0 4	1 4	1 9	2 4	2 6	1 4	2 0
1925	0 3	1 2	1 9	2 3	2 4	1 3	2 0
1926	0 3	1 1	1 9	2 2	2 2	1 3	1 9
1927	0 3	1 1	1 9	2 1	2 0	1 3	1 9
1928	0 3	1 2	2 0	2 0	1 8	1 3	1 9
1929	1 4	2 0	1 9	1 5	1 3	1 3	1 8
1930	0 4	1 7	2 1	1 8	1 3	1 3	1 8
1931	0 4	2 0	2 1	1 7	1 1	1 4	1 8
1932	0 5	2 4	2 2	1 7	0 8	1 4	1 8
1933	0 6	2 8	2 3	1 6	0 6	1 4	1 7
1934	0 8	3 3	2 4	1 5	0 4	1 4	1 7
1935	0 9	3 8	2 5	1 5	0 2	1 4	1 7
1936	1·0	4 2	2 7	1 5	0 0	1 4	1 7
1937	1 2	4 7	2 8	1 4	0 1	1 4	1 7
1938	1 4	5 0	2 9	1 4	0·2	1 4	1 7
1939	1 5	5 3	3 0	1 4	0 2	1 5	1 7

Year	*Gross National Product per capita*	*Non-farm product*	*Gross National Product*	*Non-farm capital stock (Gallman)*	*Non-farm capital stock (Kendrick)*	*Total labour force*	*Non-fa labou force*
1940	1·7	5·6	3·1	1·4	0·2	1·5	1·7
1941	5·7	3·2	1·4	0·1	1·5	1·5	1·8
1942	2·0	5·7	3·4	1·4	0·1	1·5	1·8
1943	2·2	5·7	3·5	1·4	0·4	1·5	1·8
1944	2·3	5·5	3·6	1·4	0·7	1·5	1·8
1945	2·4	5·3	3·6	1·5	1·1	1·5	1·8
1946	2·5	4·9	3·7	1·5	1·7	1·5	1·8
1947	2·5	4·5	3·8	1·6	2·2	1·5	1·8
1948	2·6	4·0	3·8	1·6	2·9	1·5	1·8
1949	2·6	3·4	3·8	1·7	3·5	1·5	1·8
1950	2·6	2·8	3·9	1·7	4·2	1·5	1·8
1951	2·6	2·0	3·9	1·8	4·8	1·4	1·9
1952	2·5	1·2	3·8	1·9	5·4	1·4	1·9
1953	2·5		3·8	2·0	5·9	1·4	1·9
1954	2·4		3·8	2·1		1·4	1·9
1955	2·2		3·7	2·1		1·3	1·9
1956	2·1		3·7	2·2		1·3	1·9
1957	2·0		3·6	2·3		1·2	1·9
1958	1·8		3·5	2·4		1·2	1·9
1959	1·6		3·4	2·4		1·2	1·8
1960	1·4		3·3	2·5		1·1	1·8
1961	1·2		3·2	2·6			1·8
1962	1·0		3·0	2·6			
1963	0·8		2·9	2·7			
1964	0·6		2·8	2·7			
1965			2·6				

MIGRATION, REMITTANCES, AND THE CASH CONSTRAINT IN AFRICAN SMALLHOLDER ECONOMIC DEVELOPMENT[1]

By ALAN RUFUS WATERS

' is now recognized that the African smallholder farmer is an efficient onomic decision-maker within the limits of the resources at his disposal id the vagaries of his environment. Unfortunately, his willingness to lapt to new opportunities makes the small farmer subject to the same verriding constraint which binds so many small businesses facing growth other societies. the inability to generate working capital fast enough to ke advantage of profitable opportunities[2] The author's experience lling agricultural chemicals to small farmers in Uganda and Kenya is led to the conclusion that working capital may be the 'ghost input' hich others have sought as an explanation for so much that appears rational in smallholder agriculture during the process of economic velopment (Johnson [38]) In the African case there are additional mplications in that credit may be unobtainable for working capital irposes, and the short-run problem may be one of physical survival ther than financial solvency

Increases in working capital must be available to the smallholder if frican agriculture is to grow at a rate which can absorb an increasing ork force in productive activity. Unfortunately, working capital is iually seen as synonymous with agricultural credit, and agricultural edit in Africa has traditionally meant small self-liquidating loans for ed amounts and on firm repayment schedules, granted only for the irchase of specific equipment or for agreed inputs required for a specific tivity[3] The availability of a growing and predictable amount of unstricted working capital is not discussed in the literature on rural credit, rhaps because our experience of rural credit has come from adminirators rather than those who have operated a small farm or business

[1] I am grateful to Professors Donald L Huddle and Robert B Ekelund, Jr, for comments an earlier draft

[2] 'Obvious as it may seem, we sometimes must remember that there is a titanic distinction tween "net profit" as opposed to the excess of cash receipts over cash disbursements' itcholl [56, p 69]) The literature on African farmers' responses to economic opportunities well known We only need cite the summary up to 1965 in Edwin Dean [17], and for a ore recent statement, Ruttan and Hayami [67]

[3] The authorities recognize the cash constraint faced by the small farmer, but they do lo about it and at best talk in terms of providing one-shot loans to enable the farm erator to 'get off the ground', i e pump-priming rather than true working capital (Vasthoff 7, pp 26, 80–1, 108]) Most advances in African agriculture have been self-financed Wilde [18, p 198]). Such credit as has become available to individual farmers has gone the larger farmers (Hayami and Ruttan [35, p 294], Turnham [74, pp 104–6])

In this article it will be argued that remittances from migrants in the high-wage sector, the cities and plantations, are a significant source of working capital for the smallholder sector. Therefore, any attempts to stimulate growth of output in the smallholder sector, and any attempts to reduce the disparity between urban and rural living standards through a shift of resources to the smallholder sector, will have to take into account the effect which such policies may have on the supply of working capital funds to agriculture.

Smallholders' cash outlays

The rapid evolution of African rural markets has reached the point where it can no longer be said that there is a non-monetary sector, except perhaps among some small and economically insignificant groups in remote areas (Mboya [**51**]) Within the active smallholder agricultural sector, barter has all but disappeared and there is widespread reliance upon the use of cash [1] Hiring of casual labour is found throughout the more productive areas of smallholder agriculture, not only for work on cash crops, but also on the traditional women's tasks of basic food production.[2] Cash expenditures can be divided into two categories direct and indirect production spending Direct production spending includes all those outlays which the farmer must make to people outside his immediate family if he is to produce a particular crop or level of output at the end of a season [3] Such outlays would include not only payments to hired labour, but also expenditure on items such as: tools, chemical fertilizers, pesticides, herbicides, livestock, fresh seed, fencing and building materials such as barbed wire and roofing sheets, as well as any service charges incurred in the crop-marketing process.

Indirect production spending can only be separated into current consumption and investment if very arbitrary definitions are introduced Family living costs, including that form of social security involved in care of the very young and the elderly, are necessary at some level for

[1] I found this to be true in the Meru district north of Mount Kenya. I was aware of the level of dependence on cash while stationed there in the early 1950s, and discovered it to have progressed even further when I returned to undertake a survey of smallholder cotton farmers in 1967 Food, for example, was moving from one area to another within the district and prices for basic staples were above those quoted in the city markets of Nairobi during parts of the year

[2] Hired labour was being used at every one of the 320 small farms which I sampled during a survey in Meru, Kenya, but the phenomenon has been noted by other observers (Bosrup [**10**, p 106], and McArthur [**52**, p 134]) An average of 50 per cent of the total labour input on the farms surveyed was hired If this widespread use of hired labour seems at variance with the concept of disguised unemployment in agriculture, it should be pointed out that labour shortages within agriculture have been consistently recorded by colonial administrators, veterinary, agricultural, and extension officials from the earliest times (McLoughlin [**53**])

[3] The season will vary with crop and location, but in agriculture there is a recognizable time period between the initiation of production and the sale of the resulting crop

the continued existence of the farm production unit. Improved health, through better and larger diets and through increased medical attention, should perhaps be called investment, but if these are available to all the farms in an area they can easily be considered short-run spending necessary to maintain the relative position of the production unit. School fees and other education costs are a significant element of smallholder cash expenditure in Africa, these could be considered investment expenditures, if they raise future productivity, but they also contain a large element required for the maintenance of the existing level of skills It is impossible to separate the components of investment from expenditure for current production. Therefore, all cash outlays by the smallholder will be treated as spending for current production, and any attempt to increase current production must entail an increase in both prior and current cash outlays.

We will define as working capital the total amount of money held at the beginning of the season for expenditure on current production and living until the proceeds from the crop become available We will also assume that the smallholder cannot expect any inflow of cash during the season. The problem faced by the smallholder now becomes part of a more general problem of working capital availability for economic development Working capital is both a constraint and the source of the ability to increase output and adopt new techniques in the future (Mosher [58, p 141]). Without an increase in working capital from some source, or a reduction in the amount of working capital required to sustain a given level of output, there cannot be growth in output over time [1]

The start and early growth of any small farm or business in the more developed nations is watched closely, and success or failure predicted, in terms of working capital availability, and this in a world where working capital may be available in the form of renewable operating credit, secured by general lien alone, from a multitude of sources (Reid [64]) The African smallholder is expected to operate on narrow margins and without access to general credit, we appear to hope that he is a better planner than his equivalent in the more developed nation, despite the acknowledged vagaries of tropical agriculture, and we appear to believe that he is more resilient in case of failure because he can fall back on some assumed supply of home-produced food which will support him and his

[1] We should note that new high-yield crops require significant increases in labour and other inputs, as has been dramatically brought out by the first results of the 'green revolution' (Johnston and Cownie [39]) Also, intensive farming in general is recognized as requiring a larger amount of working capital (Dumont [20, p 434], and Cepède [14, p 2]) And we have numerous indications of the output of cash crops rising, and cash incomes rising dramatically, with no reduction in food production at the same time when increased working capital becomes available (Taylor [73])

family if necessary.[1] Part of the belief about smallholder farmers is tied to a misunderstanding of the way in which credit actually becomes available. What little money from the official credit institutions becomes available to the smallholder farmer is tied to specific activities or to the purchase of particular equipment, it is tied also to rigid repayment schedules and does not give the farmer the flexibility which goes with the assurance of continued credit availability in the future [2]

Working capital requirements

The working capital requirements of the smallholder farmer are based upon: (*a*) the length of the production period for the crop he produces with the technology he uses, (*b*) the cost of the owned and non-owned inputs necessary to produce the crop, (*c*) the efficiency of the markets in which he buys his inputs and sells his crop, and (*d*) his subjective and highly local view of the risk and uncertainty inherent in his production process Given the diversity of local climate and soil conditions in Africa and the range of pests and diseases which exist, (*d*) may be highly significant and a factor which external credit agencies are least able to estimate with any degree of accuracy The problem facing the farmer can b discussed within a simple framework.

In the top half of Fig 1, $M_a C$ is a transformation curve for workin capital (WC) into productive inputs during a single season For simplicit we will assume that a quantity of inputs OA, perhaps as part-time famil labour or children after school hours, is available without any outlay (WC Also, we will assume the WC consists of cash alone,[3] and that th initial net endowment of WC from the previous season, OM_a, is the onl WC which will become available to the farmer until he sells his crop a the season's end The minimum WC balance which the farmer feels h must hold at all times is OM_b The slope of the transformation curv

[1] The author's experience in the East and West Mengo districts of Buganda, in Ugand during the late 1950s indicated that the trading companies believed that a small deal required a margin of 30 to 40 per cent if he was to survive, at the same time the admi istration and agricultural authorities were talking about increasing taxes and the coff cess because the smallholders were estimated to be earning in excess of 20 per cent on the operations

[2] It is interesting to note that 'credit worthiness' may be a causal factor in decidin which farmers are able to use more and higher inputs to produce the higher output which then ascribed to better 'management' (Clough [15, pp 3–5]) Clough also points to th emphasis placed on long- and medium-term equipment and mortgage credit, while survey indicate *that short-term credit may well be the vital constraint limiting the timing*, amoun and quality of inputs used, in this connection also see de Wilde [18, pp 198–207] Th concentration on repayment schedules, rather than continually expanding credit avai ability, and the search for some external measure of 'credit worthiness', are mentioned i many places (Makings [47], Jakhade [37], Binhammer [9, p 4], and Hunter [36, pp 99 100]).

[3] This is not unrealistic in a society where credit is not widespread, banking is not factor, and wages and supplies are paid for in cash.

ɘpresents the price of inputs in terms of WC and the farmer will continue ɔ apply inputs, at the appropriate time during the season, up to a total f $M_b B$ The price of purchased inputs is assumed to remain constant, ıeir supply curve being completely elastic at that price for our individual nallholder during the season. This assumption can be relaxed to take

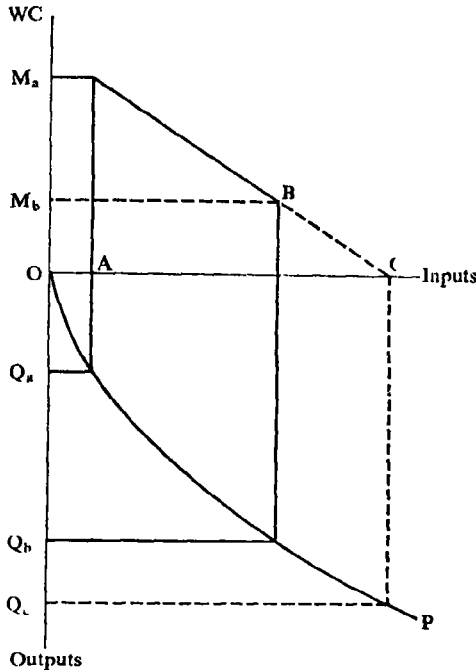



FIG 1

ccount of the observed rise in labour costs during certain periods of ıtense activity on specific crops.[1] The lower half of Fig 1 represents a ypothetical production function, *OP*, which relates various quantities f the inputs, shown horizontally, to resulting outputs on the vertical axis. ll magnitudes increase from zero at the centre. Output OQ_a would result 'om the application of *OA* non-purchased inputs, and the fact that the 'ope of *OP* indicates decreasing returns throughout as more variable ıputs are applied to the given farm is not important to our argument at his point.

Given that the farmer operates within the framework of Fig 1, we

[1] For example, the coffee crop in many parts of Kenya requires intensive concentration labour for harvesting during a brief period of time In this case it has been suggested ıat there is a seasonal labour shortage, but what is being referred to is only a shortage the going, off-season, wage As with many other apparent seasonal labour shortages, ı increase in relative wages causes it to disappear

can examine the several ways in which his situation may be altered so to induce or permit him to increase his output beyond OQ_b First, we m. be able to persuade him to extend the relevant portion of the tran formation curve and reduce his minimal cash holdings below OM_b. If increases his purchases of inputs along BC, his output would increase fro OQ_b towards OQ_c. There are two ways in which this might be done · fi we could reduce the producer's estimate of time between the end of t production season and his receipt of the proceeds from the sale of h crop; second, we could reduce the producer's estimate of the general lev of uncertainty he faces [1] Reduction in processing and marketing tir comes from the development of more efficient markets, and this tak more than a single season; but in the case of certain crops it is possib to assure the farmer of payment at some fixed time after delivery to a official agent.[2] The *assurance* of speedier payment might induce t. farmer to reduce his minimum level of WC, and increase output. Redu tions in the farmer's estimate of the uncertainty he faces should occ over time as economic development proceeds and markets improv However, because the element of risk in tropical African agriculture h a highly local component,[3] it may prove difficult for the outsider government agency to develop sufficient local knowledge to influence t farmer's estimate of local conditions to a great extent. Neverthele there are several steps of a general nature which can be undertaken the short-run

The administrative cost renders widespread crop insurance impractic for African smallholder agriculture. On the other hand, guaranteed pri can in the short-run remove one uncertainty which the producer ma face The distorting effect of such minimum prices on resource allocatio may, however, more than offset any short-run gain in output. Anoth short-run solution might be to assure the farmer of the availability credit should he want it to bolster his working capital towards the end the season The inflexibility and parsimoniousness with which smallhold credit has been administered in the past makes it appear unlikely th present institutions could change sufficiently to be a believable source assurance for the future.[4]

[1] Note that we do not try to reduce *our* estimate of the uncertainty the producer fac our estimate may differ from his This requires field survey work and close contact wi the farmer (Waters [81]).

[2] This would be the case with coffee, cotton, tobacco, pyrethrum, groundnuts, coco and various other crops which are handled by marketing boards or co-operative societie

[3] For example, the variation in tropical weather on the African plateau. It can be rain heavily on one farm and parched dry next door at the same time

[4] See n 3 above Also note that even in cases where credit has been regarded as rel tively widely available, as for example under the Tea Development Authority in Ken only a small fraction of farmers have in fact been able to obtain loans (de Wilde [19, p 62

As long as non-price rationing of credit prevails and such goals as quity based upon the widest possible distribution of available funds are ccepted, and as long as the local farmer's credit 'needs' are evaluated xternally, so long will credit be ineffective for WC purposes and inefficient for many other purposes The artificially low interest rates aintained by governmental authorities with the intention of stimulating ivestment will have the reverse effect if they cause available funds to e funnelled into the urban sector and not made accessible to the smallolder farmer who may have the incentive to be highly efficient in the use f borrowed funds and also prepared to pay a market interest rate for the rivilege [1] As Wai [79] points out, we accept interest rates of 36 per cent s normal for small loans in the more credit-worthy sectors of the markets f more developed nations, but we are ill at ease with the effective rates f over 50 per cent which prevail in the more risky unregulated credit arkets of the less developed countries

Pricing policies already in existence have one potential impact on WC oldings of African farmers Where input prices are subsidized the ransformation curve is rotated as in the upper half of Fig 2, and as a esult the quantity of inputs used will shift outwards from B towards B' ith a resulting increase in output from Q^p_b towards $Q^p_{b'}$. If the farmer els that he can safely hold a lower minimum WC balance as a result of he lower price per input, the effect of input subsidies may be even more ronounced and output may rise beyond $OQ^p_{b'}$. It is, however, unlikely hat this will occur because minimum WC balances will be determined y conditions and expectations at the season's end, when much of the WC epresents support for the family until the proceeds are received from the ale of the crop Some agricultural inputs are already subsidized in all he African nations, and it is doubtful if very much more can be done in his respect in the short run [2] We do not know if the effect of artificially igh urban wages acts to keep rural wages for less skilled labour lower han would otherwise be the case, given the relative sizes of the two ectors for most African nations it would seem unlikely that the effect is ignificant

An increase in output could also be obtained through a once-only icrease in the farmer's initial holding of WC This can be seen in Fig 2 s an upward and parallel shift of the transformation curve from, say,

[1] On the implications of financial dualism, see Myint [59]

[2] It is to be hoped that input prices will fall as world fertilizer and chemical output creases, but there is no assurance that demand will not catch up with supply as new gh-yielding crops replace present varieties which require less chemical inputs per given nd unit Also, the demand for farm output will probably rise rapidly in Africa as both comes and population continue to grow However, input subsidies appear to have been significant factor in easing WC shortages in the successful case of Japanese agricultural evelopment, as indicated by Ranis [63, p 40]

can examine the sever[al] ...
to induce or permit ...
be able to persua[de] ...
formation curve ...
increases his p...
OQ_b towards ...
we could ...
product...
crop ...
of ...

...n output of an exogenous increase in
...y certainty unless we know what pro-
... will go into non-WC investment uses
...le for major livestock or equipment
increasing access to cash is likely to

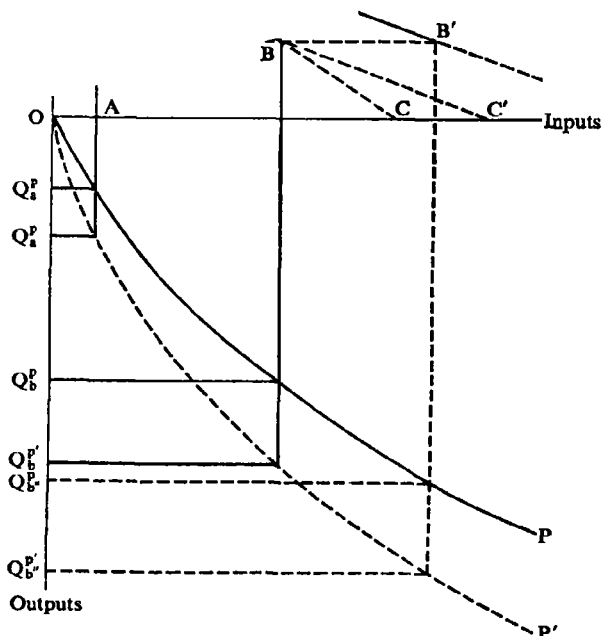



Fig 2

The ideal might be an outright transfer of resources, as represented by cash, from the urban and industrial sectors to more directly productive activities in agriculture The implied reduction in urban spending is, however, unlikely to be politically feasible in Africa. Inflation will result if increased agricultural WC is financed by permitting the money supply to grow The institutional arrangements in Africa carry a distinct urban bias, urban groups, in particular government employees, are able to obtain increases in money salaries and wages in response to rises in cost-of-living indexes which are weighted heavily in favour of those non-agricultural goods and specialized services which are used and consumed

dominantly in the urban rather than the small farming sector There-
e, because the supply of many non-agricultural goods and specialized
vices will tend to be relatively price inelastic, wages and salaries will
e disproportionately in the non-smallholder sector. The effect will be
urther transfer of wealth in the opposite direction to that for which we
arguing. To the extent that agricultural prices are permitted to rise,
d it should be noted that there is a tendency for them to be controlled
the non-smallholder sector of many African nations, real WC may
pear to increase However, real WC will increase less than the initial
netary change because of the increased relative prices of non-farm
uts. Also, inflation may partially offset increases in monetary WC where
ignificant portion of agricultural output is sold on foreign markets To
extent that inflation does result in a redistribution of wealth in society,
ay from less productive activities, it may increase real WC. However,
only way in which inflation resulting from monetary expansion, and
t directly connected to increased agricultural WC, could benefit the
mers' WC position would be if farm prices were rising faster than labour
sts, i e by a reduction in real wages [1] Where labour is the single most
portant agricultural input, and where an overwhelming proportion of
total labour force is employed in agriculture, a reduction in real wages
uld hardly be seen as an appropriate means to increase WC in a society
ich is not efficiently totalitarian

A further means to increased effective WC holdings lies in improving
e technological efficiency with which the smallholder can convert a
en quantity of inputs into output An example of this can be demon-
ated as a downward shift in the production function from OP to OP'
the lower half of Fig 2 However, the specific degree to which output
n be improved depends upon the slope of the old production function,
slope and location of the new production function, and the ability to
ld input quality constant [2] If a new production technique requires
re expensive inputs of a higher quality than before, it may not be taken
if WC requirements cannot be met (de Wilde [18, p 198], and Baker,
inghas, and Meyers [5]) This may well happen when new high-yield
ps are introduced But the introduction of new production techniques
h as interplanting and altered scheduling of farm activities, or even the
roduction of new crops or crop diversification, is unlikely to produce

See Bottomley [11] on the problems of inflation in rural areas

We have, for simplicity, looked at only one of the possible adjustments in the production
ction that due to uniform technical change Following Harry G Johnson's argument
t such geometry is a highly flexible tool, 'which can achieve significant results using only
litative restrictions on the shapes of the relevant behaviour functions and can explore
blems in the large as well as in the small' (Johnson [38A, pp 9-10]), it would be relatively
y to develop an argument for several other specific adjustments and examine their
come

change if unaccompanied by increased WC The present discounted valı of any increased future returns which the farmer may expect from son new activity will be based upon the increased risk due to change itse We have argued that risk and uncertainty are likely to be larger compo ents of cost in African smallholder agriculture than elsewhere. Therefor any change, even one which an outsider feels is bound to succeed, such a move away from monoculture, must require an increase in WC or compensating reduction in the cost of other inputs The benefits of ne techniques or crops, or diversification which will raise *future* revenue ar lower *future* costs, comes only in the seasons after their introduction hı been made possible by increased WC.

If the WC constraint cannot be relieved, due to existing institution arrangements, there may be a case for moving to larger production un with access to greater capital resources rather than trying to persua existing farmers to take risks which their local knowledge tells them t avoid But such a second-best solution is a poor alternative to attackir credit market imperfections and other institutional rigidities It will als be important, however, to test new crops and techniques over a range c possible farm sizes and then publicize that outcome which appears t put the least strain on the existing supply of WC (Sánchez [**68**])

Migration and wage-sector dualism

The lack of WC with which to take advantage of productive opportur ities in the smallholder agricultural sector may be a negative incentive a push factor, impelling migration to the cities The existence has bee recognized of a real wage differential between the rural and predominantl smallholder sector of most African economies and the government influenced sector, which for convenience we will call urban [1] Urbaı unemployment exists on a growing scale and is seen as a problem o resources misallocated or redundant African urban populations ar predominantly male and in their younger working years, and the anthropo logists tell us that the African cities are unusual in that the migrants who move to them from the rural areas feel that their stay is temporary an they continue to maintain strong ties with their homes even though they may remain in the city for many years Furthermore, the indications are that the proportion of the urban population which can be considere transient does not appear to be declining with time [2] Despite visibl

[1] For one of many good descriptions of the situation see Berg [**7**, pp 158–208] On the equivalent disparity in the provision of social services to the two sectors see Shefiel [**69**, p 28] The problem has been observed on both sides of the continent and in both town and cities (Roussel [**66**])

[2] The evidence for the size, pervasiveness, and continuing nature of transient rura migration to and from the towns and cities can be found in a wide range of studies in several disciplines For examples from throughout the continent at various points in time see

ıemployment ahead of them the migrants continue to pour ınto the ties, and to the fragile governments of the new Afrıcan natıons they ·present a danger far beyond their economıc sıgnıficance (Lee [42]). ıttle ıs known about the magnıtude of the stream of mıgrants, except ıat the pool of urban unemployed ıs believed to be growıng, there ıs ven less ınformation about the reverse flow to the rural sector, although ıch a flow is known to exıst; and there ıs practıcally no ınformatıon at all pout the size and nature of the flow of remıttances to the rural sector, though we have ındıcations that ıt ıs substantial [1]

The exıstence of a contınued urban-rural real wage dıfferentıal, ın ccess of anythıng required by dıfferıng margınal productıvıtıes of labour the two sectors, can be expected to contınue ınto the foreseeable ıture, the potential solutıons are unlıkely to be serıously trıed It ıs oubtful that any present African government could survıve the attempt ı reduce real urban wages dırectly (Berry [8]) Therefore, ıt has been ıggested that migration could be physıcally controlled, perhaps by bour laws and a pass system such as presently exıst ın South Afrıca ıd Tanzanıa. The two sectors would be ınsulated from each other untıl ıe growth ın city employment had absorbed the exısting unemployed, ıd ın the process created a rısing demand for the products of the rural ector, thus spreadıng prosperıty and reducing the ıncentıve to mıgrate Harrıs and Todaro [33] [34]) But thıs alternatıve is unlıkely to work ıf he solutıon is to take tıme and the ıncentıve to migrate remaıns hıgh ın he short-run, the more lıkely outcome ıs frustratıon and socıal upheaval n the face of growıng repressıon and control The thırd approach suggests hat the emphasıs be placed upon agrıculture, and that ın the future conomıc resources be swıtched to the rural sector and away from the ıtıes, and that the urban rate of growth be held constant, or ıf thıs ıs not olıtıcally feasıble, only be permıtted to grow at a slower rate untıl the ural sector has caught up [2] Thıs alternatıve sounds attractive, though ıt oo offers a long-run solutıon to a short-run problem, and requıres that he urban sector gıve up potentıal wealth.

Even thıs lımıted goal may be hard for a government to achıeve ıf ıts

ullıver [29], Elkan [23], Mayer [50], Panofsky [60], Spengler [71], Kuper [75], Elkan [24], ohen [16], Gugler [27], Hance [32], and Kımmerlıng [41]

[1] It may be argued that exceptıonal cırcumstances account for the great stream of remıttances from South Afrıca, but we are begınnıng to dıscover that ıt ıs not unıque There ıs so a sıgnıficant flow of remittances from all the cıtıes and towns of sub-Saharan Afrıca or some ındıcatıon of the pattern of remıttances see Ardner, Ardner, and Warmıngton , pp 183–8], Van Velson [76], Baeck [4], Little [45, pp 21–9], Plotnıcov [62, 296–9], eıstner [43], Beals, Levy, and Moses [6], Robson [65, pp 263, 267], Caldwell [13, pp 152–], Kılby [40, pp. 203–6], Gutkınd [31], Mıracle and Berry [57], and Anthony and Uchendu]

[2] Thıs approach has been wıdely advocated For a good recent statement of the argument e Abercrombıe [1]

power base and cultural roots are in and of the city (Waters [80]). ever, the situation exists, and it has several positive aspects of whi may not have been fully aware We need to know considerably about what is actually occurring as economic development procee the smallholder and rural sector of Africa, but there are already indica of certain positive and self-correcting forces at work within the urban dichotomy We should begin by noting that most of the mig to the cities appear to come from the more prosperous and rapidly gr parts of the rural sector, rather than from the stagnating and/or cro areas where the people are at the greatest disadvantage in terms of l standards.[1]

Remittances as a source of WC

We will now develop the connection between the need for addit WC in the process of smallholder development, and the stream of r tances flowing from the urban to the rural sector It is important, how to recognize that there are many motives for migration to the cit that the young men who do so in order to escape from family const and responsibility at the age of revolt are not part of our problem [2] men will migrate for a single clear reason We will only suggest tha desire to break a WC constraint is one significant element in the dec but we will assume for ease of analysis that the other elements re constant. If a smallholder in a growing area of Africa is faced w shortage of WC, and an injection of funds would enable him to inc his crop in several seasons to the point where he could take advanta new techniques and crops, he may leave the farm to earn the req funds [3] Alternatively, the rate of return to additional WC funds cou sufficiently high to induce migration even under existing circumsta One factor we should mention is the relative attractiveness of rural

[1] This observation is only true of migration to the cities I have observed the f migrants into Mengo, Masaka, and the Mubende areas of Uganda, from the poorer of the country, and how this migration was accompanied by a smaller but sign migration from these wealthier areas of Buganda into the city of Kampala Distan been recognized as a deterrent to migration, but the outlying areas have also tended the poorest parts of most African nations See Richards [21], Beals, Levy, and [6, p 484], and Manners [48, pp 345–6].

[2] It is likely that a large number of the urban unemployed represent this escaping Having been an employer in Nairobi and Kampala I do not recognize the Todaro assum that the longer a city youth is out of work the more likely he is to get a job From m experience this is a complete inversion

[3] Many examples of failure with new crops and methods in Africa could be explai terms of the poor-risk evaluator and the gambler, rather than the real innovator, bei one to take them up The better farmer might not be recognizable by external auth if he eschewed the trappings of 'good farming', and he might decide that his WC con precluded the new crop or technique

It is suggested that the drudgery and ugliness of African rural life drives people to the bright and exciting cities. There is evidence, from Beals, Levy, and Moses [6, pp 186–7] and Kimmerling [41, p 483], that this is not so.[1] We should also point out that labour market imperfections, such as those created by minimum wage laws and the political power of civil service groups or other trades unions, may limit the type of migrant to the young adult of prime working age On the other hand, the hired labour which remittances enable the small farm to use will consist of those who are tied to the farm area for a variety of reasons youth or old age, local responsibilities for which no cash flow can compensate, pure inability to migrate because of failure to raise the cost of transportation and initial living expenses, as well as those who have a strong psychological preference for their native region. Our argument, however, is fundamentally positive incentives to migrate are stronger for some individuals than for others, and incentives are likely to be very strong for the small farmer facing a WC constraint

For our purposes, the inducement for the smallholder to migrate lies in the net return which he can obtain through the use of remittances as WC on his farm The observable flow of remittances consists of several components, and the timing of increments to WC may be as significant as the magnitude of the funds involved. For a given production season, the flow of remittances will depend upon all of the human emotions, such as charity, pride, and family loyalty, but we will collect these under a single heading and say that they represent a disturbance premium necessary to induce the migrant to move if other factors are equal in both sectors This permits us to examine the main part of the remittance flow in colder terms (Gugler [27]). For the migrant to seek urban work, the net return which he derives from the use of additional and appropriately timed units of WC must exceed the gross wage available to him off the farm, less the disturbance premium and two other elements First, he must deduct the difference between the cost of supporting himself on the farm and in the urban setting, second, he must deduct the cost of replacing his own forgone labour on the farm Therefore, the absolute difference between urban and rural labour earnings does not alone determine the type of migration we are discussing, longer-run returns to farm activities, the cost of purchased inputs for the farm, the relative cost of living in both

[1] I used the opportunity presented by a major survey of coffee growing smallholder farmers in the Meru district of Kenya to ask numerous additional questions about migration I also questioned co-operative society personnel and urban employees of the coffee mill. The results cannot be offered as a formal test of the hypothesis that, wage and salary differentials apart, the rural small farm is more attractive than the city to the majority of Africans, but I was convinced that this is so Also see Peil [61, p 149], ' "the lure of the town' is a factor in the decision of some young men to leave home, but it is certainly much less important than the desire for wage employment '

sectors, and the disturbance premium necessary to compensate for ı
and hardship are the more basic determinants.

The regular inflow of funds during the season will increase the
investment if it causes lower average WC balances to be held an
quently permits the farmer to use a larger portion of the procee
the sale of his crop on non-WC uses. However, average WC bala
unlikely to decline by the full amount of the average inflow of rem
if the migrant is also the main decision-maker on the farm and t
the managerial cost of his absence, which could be included in the
ance premium, were to rise towards the end of the season when prc
decisions are more crucial On the other hand, the assurance of
inflow of funds may permit lower end-of-season WC balances
period between the harvest and receipt of the resulting incon
timing of remittances may be particularly important for areas w
highly specialized in the production of some cash crop, or wl
unsuitable for the production of basic foods (Gulliver [**30**, pp.
Finally, there are two potentially dangerous effects which can st
eventual dependence upon a flow of remittances once it has bee
way for some time If remittances are used collectively, as is the c
the urban associations of certain areas in West Africa, there m
tendency to establish new plantations in the home district v
available funds (Smock [**70**]). Similarly, the smaller holdings te
the more intensively farmed with purchased inputs and will t
require proportionately more WC (Massell [**49**]) In both the
there can develop a reliance on continued flows of WC to su
relatively large fixed cost component in the production process
turn creates a situation of vulnerability to changes in governmer
with respect to the urban-rural dichotomy, and also vulneral
changes in output prices. For the moment, however, these danger
relatively minor in the light of the over-all need for more WC.

Conclusion

Remittances, flowing in the opposite direction to the stream of n
have a dual impact on the basic dichotomy which exists between t
and urban sectors of African nations. There is an immediate distri
and welfare effect as the remittances permit a higher level of
consumption in rural areas while reducing it in the cities, and t
longer-run distributional effect via the increased availability of
capital to the growing smallholder agricultural sector The lo
effect has not been studied by economists as yet, perhaps becau
almost total lack of data.[1] '. . . painfully little of a specific nature ı

[1] The effect of remittances on the level of disguised unemployment which is
exist in some less-developed areas is brought into the picture by Mehmet [**54**].

about the process of capital formation and the need for reproducible and working capital in agricultural development' (Lewis [44, p 461])

Two distinct empirical economic studies are needed for all major African cities.[1] First, studies of the sources of remittances in the cities and the size of remittance flows to various specific areas and sectors of the economy, second, studies of the destination of remittance flows and the uses to which funds are put in differing kinds of economic activity Motivational and behavioural studies will have a relatively lower priority for economists at this early stage than surveys and fieldwork to discover the size of cash balances and the timing and magnitude of cash flows

We also need to know much more about the nature of the working capital constraint faced by the African smallholder It has been suggested that a shortage of working capital may account for the observed under-utilization of capital in the agricultural as well as the industrial, sectors of less developed economies (Little [46, p 12]) It has also been suggested that small farmers are forced to cultivate low value and rapid turnover crops due to their inability to finance the higher cost inputs and the longer production period required of crops which will eventually yield a much higher return [2]

In spite of our recognized lack of knowledge, there are several general policy suggestions, all already offered for a variety of other good reasons, which can be made It would be advisable not to pursue taxation policies which may cause a direct drain on smallholder working capital balances This may mean that such cash taxes as the graduated poll tax in Kenya should be reconsidered and replaced by an alternative until more is known about its impact on agricultural output There may also be grounds for permitting a higher rate of inflation if the positive effect in the smallholder sector more than offsets the negative effects elsewhere in the economy More important, if somewhat longer-term in effect, would be policies to broaden the credit market by permitting interest rates to rise so that capital would become more available outside the urban sector, and creating new types of negotiable credit instruments and the market institutions for discounting and rediscounting them Higher interest rates will offer an inducement to the lenders who now consider small farming too risky even to approach, and would also permit the establishment of a new layer of small and local lending institutions with

[1] The sociologists and anthropologists are still involved in seeking a framework within which to study the African city, they are still at the taxonomic stage (Epstein [25])

[2] See Vyas [78] on the efficiency of small farmers elsewhere than Africa We have been aware for a long time that African farmers know how to be good conservationists and eliminate such disinvesting forces as soil erosion if it is in their interests to do so (Stamp [72]). Again, the problem of their failure to undertake long-run investments may be explained in terms of their overriding need for WC in the short-run (Vasthoff [77, p 15])

access to local knowledge and information in the rural areas. The righ of foreclosure will have to be enforced if credit is to be rationed in term of economic efficiency, and that in turn requires expanding markets fo all the assets used in farming (Bottomley [12]) Finally, the over-al policy for economic growth in Africa must focus on agriculture; and thi means smallholder agriculture.[1] Because smallholder agriculture is wide spread and consists of a multitude of small production units, marke efficiency and market flexibility are particularly susceptible to transporta tion constraints The transportation systems of Africa are city oriente and subject to numerous arbitrary licensing laws and constraints whicl cause a bias in the type of equipment used; this must change, and i requires a reduction rather than an increase in current operating costs t bring about such change (Meyer [55])

Evolution and growth will themselves create the conditions which ar needed to channel more working capital to the smallholders in agriculture Urban remittances provide a source of funds in the interim, and therefor speed the process of growth. In this respect remittances are a factor i ameliorating the abrasive process of readjustment between city and rura living standards which will have to occur in the future (Eicher [22])

Texas A and M. University

REFERENCES

1. ABERCROMBIE, K. C., 'Fiscal policy and agricultural employment in developing countries', *Monthly Bulletin of Agricultural Economics and Statistics*, vol. 20 no. 3, Mar. 1971, pp. 1–7.
2. ANTHONY, KENNETH R M., and UCHENDU, VICTOR C., 'Agricultural change in Mazabuka District, Zambia', *Food Research Institute Studies*, vol ix, no 3 1970, pp. 215-67.
3. ARDENER, EDWIN, ARDENER, SHIRLEY, and WARMINGTON, W A, 'Saving remittances home and debts', in *Plantation and Village in the Cameroons, Some Economic and Social Studies*, London, Oxford University Press, 1960.
4. BAECK, L., 'An expenditure study of the Congolese évolués of Leopoldville Belgian Congo', in Aidan Southall, ed : *Social Change in Modern Africa*, London Oxford University Press, 1961, pp. 159–81.
5. BAKER, RANDOLPH, MAHAR, MANGHAS, and MEYERS, WILLIAM H., "The probable impact of the seed-fertilizer revolution on grain production and on farm labor requirements', a paper presented at the Conference on Strategies for Agricultural Development in the 1970s, Stanford University, Dec. 1970.
6. BEALS, RALPH E., LEVY, MILDRED B., and MOSES, LEON N., 'Rationality and migration in Ghana', *The Review of Economics and Statistics*, vol. xliv, no. 4 Nov. 1967, pp. 480–6

[1] The question of economies of scale in agriculture is not being opened, though it has been questioned as an assumption for a considerable time (Frankel [26, p 149]).

7. BERG, ELLIOTT J., 'Major issues of wage policy in Africa', in A. M. Ross, ed.: *Industrial Relations in Economic Development*, New York, Macmillan, 1968.
8. BERRY, SARA S., 'Economic development with surplus labour: further complications suggested by contemporary African experience', *Oxford Economic Papers* (New Series), vol. 22, no. 2, 1970, pp. 275–82.
9. BINHAMMER, H. H., 'Institutional arrangements for supplying credit and finance to the rural sector of the economy in Tanzania', Economic Research Bureau Paper 68. 17, University College, Dar es Salaam, 1968
10. BOSERUP, ESTER, *The Conditions of Agricultural Growth, The Economics of Agrarian Change Under Population Pressure*, Chicago, Aldine, 1965.
11. BOTTOMLEY, ANTHONY, 'A monetary strategy for underdeveloped rural areas', *Journal of Agricultural Economics*, vol xvii, no 2, Sept 1966, pp 139–49
12. —— 'The premium for risk as a determinant of interest rates in underdeveloped rural areas', *Quarterly Journal of Economics*, vol. lxxvi, no. 4, Nov. 1963, pp. 637–47.
13. CALDWELL, JOHN C, *African Rural-Urban Migration, The Movement to Ghana's Towns*, New York, Columbia University, 1969.
14. CÉPÈDE, MICHAEL, 'The green revolution and employment', *International Labor Review*, vol. 105, no 1, Jan 1952, pp 1–8.
15. CLOUGH, R. H., 'Recent experience with farm management surveys in Kenya', a paper presented at the East African Agricultural Economics Society Conference, Dar es Salaam, Apr. 1970.
16. COHEN, ABNER, *Custom and Politics in Urban Africa*, Berkeley, University of California, 1969
17. DEAN, EDWIN, *The Supply Responses of African Farmers*, Amsterdam, North-Holland, 1966.
18. DE WILDE, JOHN C, *Experiences with Agricultural Development in Tropical Africa*, vol. i, 'The Synthesis', Baltimore, Johns Hopkins, 1967
19. —— *Experiences with Agricultural Development in Tropical Africa*, vol ii, 'The Case Studies', Baltimore, Johns Hopkins, 1967
20. DUMONT, RENÉ, *Types of Rural Economy*, London, Methuen, 1957.
21. *Economic Development and Tribal Change A Study of Immigrant Labour in Buganda*, Audrey I. Richards, ed., Cambridge, W. Heffer, 1952.
22. EICHER, CARL, 'Tackling Africa's unemployment problems', *Africa Report*, vol. 16, no. 1, Jan. 1971, pp. 30–3.
23. ELKAN, WALTER, *Migrants and Proletarians Urban Labour in the Economic Development of Uganda*, London, Oxford University Press, 1960.
24. —— 'Circular migration and the growth of towns in East Africa', *International Labor Review*, vol. xcvi, no 6, Dec. 1967, pp 581–9.
25. EPSTEIN, A. L., 'Urbanization and social change in Africa', in Gerald Breese, ed.: *The City in the Newly Developing Countries*, Englewood Cliffs, New Jersey, Prentice Hall, 1969, pp. 246–83.
26. FRANKEL, S. H., *The Economic Impact on Underdeveloped Societies*, Oxford, Basil Blackwell, 1952.
27. GUGLER, JOSEF, 'On the theory of rural-urban migration, the case of sub-Saharan Africa', in J. A. Jackson, ed *Migration*, Cambridge University Press, 1969, pp. 134–55
28. —— 'The impact of labor migration on society and economy in sub-Saharan Africa. Empirical findings and theoretical considerations', *African Social Research*, no. 6, Dec. 1968, pp. 463–86.
29. GULLIVER, P. H., 'Incentives in labour migration', *Human Organization*, vol xix, no. 3, Fall 1960, pp. 159–63.
30. —— 'Labour migration in a rural economy', *East African Studies*, no 6 Reprinted in Edith H. Whetham and Jean I Currie, eds *Readings in the Applied*

Economics of Africa, vol. 1, 'Micro-Economics', Cambridge University Press, 1967.

31. GUTKIND, PETER C. W., 'African unionism, mobility and the social network', in Gerald Breese, ed.: *The City in the Newly Developing Countries*, Englewood Cliffs, New Jersey, Prentice-Hall, 1969, pp. 389–400.
32. HANCE, WILLIAM A., *Population, Migration and Urbanization in Africa*, New York, Columbia University, 1970.
33. HARRIS, JOHN R., and TODARO, MICHAEL P., 'Wages, industrial employment and labor productivity: the Kenyan experience', *East African Economic Review* (New Series), vol. 1, June 1969, pp 29–46.
34. —— ——'Migration, unemployment and development: a two-sector analysis', *American Economic Review*, vol lx, no. 1, Mar 1970, pp. 126–42.
35. HAYAMI, YUJIRO, and RUTTAN, VERNON W., *Agricultural Development An International Perspective*, Baltimore, Johns Hopkins, 1971.
36. HUNTER, GUY, *The Administration of Agricultural Development, Lessons from India*, Ames, Iowa State University, 1971.
37. JAKHADE, V. M., 'Small farmers and cooperative credit', *Problems of Small Farmers*, Bombay, Indian Society of Agricultural Economics, 1967, pp. 83–92.
38. JOHNSON, GLEN L., 'A note on non-conventional inputs and conventional production functions', in Carl K. Eicher and Lawrence W. Witt, eds.: *Agriculture in Economic Development*, New York, McGraw-Hill, 1964, pp. 120–4.
38A. JOHNSON, HARRY G., *The Two-Sector Model of General Equilibrium*, Chicago, Aldine-Atherton, 1971.
39. JOHNSTON, BRUCE F., and COWNIE, JOHN, 'The seed-fertilizer revolution and the labor force absorption problem', *American Economic Review*, vol. lix, no. 4, pt. 1. Sept. 1969, pp. 569–82.
40. KILBY, PETER, *Industrialization in an Open Economy Niger 1945–66*, Cambridge University Press, 1969
41. KIMMERLING, BARUCK, 'Subsistence crops, cash crops, and urbanization: some materials from Ghana, Uganda, and the Ivory Coast', *Rural Sociology*, vol. 36, no. 4, Dec. 1971, pp. 471–87.
42. LEE, EVERETT S., 'A theory of migration', *Demography*, vol iii, no 1, 1966, pp. 47–57.
43. LEISTNER, G. M. E., 'Foreign Bantu workers in South Africa their present position in the economy', *The South African Journal of Economics*, vol. xxxv, no 1, Mar. 1967, pp. 30–56.
44. LEWIS, STEPHEN R., Jr., 'Agricultural taxation in a developing economy', in Herman Southworth and Bruce F. Johnston, eds.: *Agricultural Development and Economic Growth*, Ithaca, New York, Cornell University, chpt 12, 1967.
45. LITTLE, KENNETH, *West African Urbanization*, Cambridge University Press, 1965.
46. LITTLE, I. M. D., 'The influence of economic policy in less developed countries on the capital intensity of investment, and growth of employment', a paper presented at the Conference on Strategies for Agricultural Development in the 1970s, Stanford University, Dec. 1971.
47. MAKINGS, S M., *Agricultural Problems of Developing Countries in Africa*, Lusaka, Oxford University Press, 1967
48. MANNERS, ROBERT A., 'The Kipsigis of Kenya. culture change in a "Model" East African Tribe', in Julian H Steward, ed.: *Contemporary Change in Traditional Societies*, Urbana, University of Illinois, 1967.
49. MASSELL, BENTON F., 'Farm management in peasant agriculture: an empirical study', *Food Research Institute Studies*, vol. vii, no. 2, 1967, pp. 205–15.
50. MAYER, P., 'Migrancy and the study of Africans in towns', *American Anthropologist*, vol. lxiv, no. 3, pt. 1, June 1962, pp. 576–92.

. MBOYA, TOM J., 'The impact of modern institutions on the East African', in P. H. Gulliver, ed.: *Tradition and Transition in East Africa*, Berkeley, University of California, 1969, pp. 89–103.
. MCARTHUR, J. D., 'Some thoughts on future trends in farm employment in Kenya', in James R. Sheffield, ed.: *Education, Employment and Rural Development*, Nairobi, East African Publishing House, 1967, pp. 122–40
. MCLOUGHLIN, PETER F. M., 'The need for a "full employment" and not a "disguised unemployment" assumption in African development theorizing', *Zeitschrift fur Nationalokonomie*, Band xxii, Heft 4, Dec 1962, pp 361 7
. MEHMET, OZAY, 'A note on the disguised unemployment hypothesis', *Economia Internazionale*, vol xxiv, no 1, Feb. 1971, pp 113 16
. MEYER, JOHN R., 'Transport technologies for developing countries', *American Economic Review*, vol. lvi, no. 2, May 1966, pp 81–91
. MITCHELL, WILLIAM E., 'Cash forecasting the four methods compared', in Edward J. Mock, ed *Readings in Financial Management*, Scranton, Pennsylvania, International Textbook, 1964.
. MIRACLE, MARVIN P., and BERRY, SARA S., 'Migrant labour and economic development', *Oxford Economic Papers*, vol 22, no. 1, Mar. 1970, pp 86–108
. MOSHER, A. T., *Getting Agriculture Moving*, New York, Praeger, 1966
. MYINT, H., 'Dualism and the internal integration of the underdeveloped economies', *Banca Nazionale Del Lavoro Quarterly Review*, no 93, June 1970, pp 128–56
. PANOFSKY, HANS E., 'Migratory labor in Africa a bibliographical note', *Journal of Modern African Studies*, vol 1, no 4, 1963, pp 521–9
. PEIL, MARGRET, *The Ghanaian Factory Worker Industrial Man in Africa*, Cambridge University Press, 1972.
. PLOTNICOV, LEONARD, *Strangers in the City, Urban Man in Jos, Nigeria*, University of Pittsburgh, 1967
. RANIS, GUSTAV, 'The financing of Japanese economic development', in Bruce F. Johnston and Hiromitsu, eds *Agriculture and Economic Growth Japan's Experience*, New Jersey, Princeton University, 1970
. REID, IAN G., 'Accounting for capital in the farm business', *Journal of Agricultural Economics*, vol. xix, no 2, May 1968, pp. 177 91.
. ROBSON, PETER, *Economic Integration in Africa*, London, Allen & Unwin, 1968.
. ROUSSEL, LOUIS, 'Employment problems and policies in the Ivory Coast', *International Labor Review*, vol 104, no. 6, Dec 1971, pp 505–25.
. RUTTAN, VERNON W., and HAYAMI, YUJIRO, 'Strategies for agricultural development: the evolution of thought', a paper presented at the Conference on Strategies for Agricultural Development in the 1970s, Stanford University, Dec. 1971.
. SÁNCHEZ, LEOBARDO JIMÉNEZ, 'Strategies for increasing agricultural production from small holdings. the Puebla Project', a paper presented at the Conference on Strategies for Agricultural Development in the 1970s, Stanford University, Dec. 1970.
. SHEFFIELD, JAMES R., 'Conference conclusions', in James R. Sheffield, ed *Education, Employment and Rural Development*, Nairobi, East African Publishing House, 1967.
. SMOCK, DAVID R., 'Cultural and attitudinal factors affecting agricultural development in Eastern Nigeria', *Economic Development and Cultural Change*, vol. 18, no. 1, Oct. 1969, pp. 110–24.
. SPENGLER, J. J., 'Population movements and problems in sub-Saharan Africa', in E. A. G. Robinson, ed.: *Economic Development for Africa South of the Sahara*, New York, St. Martin's, 1964.

72. STAMP, L. DUDLEY, 'Land utilization and soil erosion in Nigeria', *The Geographical Review*, vol. xxviii, Jan. 1938, pp. 32–45.
73. TAYLOR, D. R. F., 'Changing land tenure and settlement patterns in the Fort Hall District of Kenya', *Land Economics*, vol. xl, no. 2, May 1964, pp. 234–7.
74. TURNHAM, DAVID, *The Employment Problem in Less-Developed Countries*, Paris, O.E.C.D., 1971.
75. *Urbanization and Migration in West Africa*, Hilda Kuper, ed., Berkeley, University of California, 1965.
76. VAN VELSEN, J., 'Labor migration as a positive factor in the continuity of Tonga tribal society', *Economic Development and Cultural Change*, vol. 8, no. 3, Apr. 1960, pp. 256–78.
77. VASTHOFF, JOSEF, *Small Farm Credit and Development, Some Experiences in East Africa with Special Reference to Kenya*, Munich, Weltforum Verlag, 1968.
78. VYAS, V. S., 'Economic efficiency on small farms of central Gujarat', in *Problems of Small Farmers*, Bombay, Indian Society of Agricultural Economics, 1967, pp. 60–75.
79. WAI, U. TUN, 'Interest rates outside the organized money markets of underdeveloped countries', *IMF Staff Papers*, vol. vi, no. 1, Nov. 1957, pp. 80–142.
80. WATERS, ALAN RUFUS, 'A behavioral model of pan-African disintegration', *African Studies Review*, vol. xiii, no. 3, Dec. 1970, pp. 415–33.
81. ——'Collecting economic information in rural Africa', a paper presented at the Annual Conference of the Southern Economic Association, Miami Beach, Nov. 1971.

ARE THERE REAL LIMITS TO GROWTH?— A REPLY TO BECKERMAN

By LOWELL S. BROWN, LEONARDO CASTILLEJO, H. F JONES, T. W B. KIBBLE, *and* M ROWAN-ROBINSON

In his recent inaugural lecture[1] Professor Beckerman threw out a challenge to scientists to state their views on *The Limits to Growth*,[2] the computer-based study by Meadows *et al.* sponsored by the Club of Rome. 'Why', he asks, 'have the professional scientists, with a few notable exceptions, such as Kenneth Mellanby, kept fairly quiet about it all, or have limited themselves to relatively mealy-mouthed criticisms?'

It is true that there have been few outright condemnations of the book by scientists, and one may well ask why. In this note we take up the challenge to express our views in the hope of shedding some light on the differences in outlook between scientists and economists which produce such contrasting responses

Many scientists have mixed reactions to *The Limits to Growth* [3] In view of the lack of easily available information about the details of the model, they have generally preferred to reserve final judgement Various specialists have now examined these details, and their considered criticisms are beginning to appear [4] But although *The Limits to Growth* may be shown to be wrong in some of its assumptions, and may perhaps be unduly pretentious, it has been of great value at least in stimulating debate on a very real issue, one which many economists have been reluctant to confront Certainly none of the book's shortcomings can justify the arrogant complacency of Beckerman's closing remarks [5]

The essence of the environmentalist case is that man is growing in numbers, technical ability, and demand for natural resources to such a degree that his activities and the pollution they produce are ceasing to be a small perturbation of the global ecosystem In these circumstances there is a clear prima-facie case for the possibility of overshoot and collapse. The situation may not be as clear-cut as *The Limits to Growth* makes out,

[1] W. Beckerman, 'Economists, scientists, and environmental catastrophe', *Oxford Economic Papers*, vol 24, no. 3, pp 327–44, Nov. 1972

[2] Meadows, Meadows, Randers, and Behrens, *The Limits to Growth* (London, Earth Island, 1972).

[3] For two such views see Garrett Hardin, 'We live on a spaceship', *Bulletin of the Atomic Scientists*, vol. 28, no. 9, pp. 23–5, Nov. 1972, and R S. Berry, 'Reflections on "The Limits to Growth"', ibid., pp 25–7

[4] For example, Oerlemans, Telling, and de Vries, *Nature*, vol 238, p 251 (4 Aug 1972). See also especially the report of the work of the Sussex group, 'Thinking about the Future', *Futures* (1973), vol. 5, nos. 1 and 2

[5] W. Beckerman, op. cit, pp. 343–4.

but the possibility cannot be dismissed as easily as Beckerman seems t imagine.

The basic flaw in Beckerman's approach, and the reason that all h curious analogies from ancient Greece or Egypt are totally irrelevant, is h failure to understand the significance of the relative orders of magnituc of human and natural effects The present situation is qualitatively ne because of the *scale* of man's activities. This is in part simply a reflection o our numbers, but also of the increasing demands of modern technology Fo instance, in 1970 the U S A consumed energy at a rate of about 10 kW pe person, as compared with about 0·3 kW in the less developed countries (an until not long ago in the more developed ones too).[1] In some areas, energ usage already amounts to a few per cent of the incident solar radiation

Our single species now consumes directly or indirectly (by eating cov that eat grass) some three per cent of all the new biomass produced b photosynthesis on the entire land surface of the earth [2] An ecological syste dominated in this way by one species has a well-recognized tendency t instability,[3] and one incidentally which cannot be countered by an obvious economic mechanism. Though we may yet learn how to avoid catastrophic collapse, we would be well advised not to push our dominanc too far

This question of scale is absolutely fundamental to any proper discussio of limits to growth For almost any single problem we can with enoug effort invent a technological solution, finding substitutes for scarce re sources, mining poorer ores, and so on (although there are some materia for which no substitutes *can* exist, such as phosphates,[4] which play a precis biochemical role) What is at issue is our ability to do all the things tha will be needed at the same time, without creating more problems than w solve. Yet nowhere in Beckerman's lecture is there any mention of sca in this sense. It is very significant that all the examples he quotes to sho that large improvements in the elimination of pollution are achievable a of a local nature—he mentions the British rivers[5] for example, but not th Mediterranean or the Baltic.[6]

[1] See Earl Cook, 'The Flow of Energy in an Industrial Society', *Scientific American*, Sep 1971, p 135

[2] See, for example, J Harte and R H Socolow, *Patient Earth*, pp 283–4 and 338 (Ne York, Holt, Rinehart, Winston, 1971) Note that this figure excludes primary productio in the oceans.

[3] C Elton, *Animal Ecology*, 3rd edn (London Methuen, 1950) See also Goel, Maitr and Montroll, 'On the Volterra and other nonlinear models of interacting populations', *Re Mod Phys* , vol 43, p 231–76 (Apr 1971), especially pp 256–8

[4] See A Ehrlich and P Ehrlich, *Population, Resources, Environment* (San Francisco Freeman, 1970)

[5] W. Beckerman, op. cit., p. 339.

[6] See the *First Report* of the *Royal Commission on Environmental Pollution* (London H M.S O , 1971), p 25

An intimately related point concerns the *time scale* on which we have to work. Perhaps given unlimited time we could find solutions to all our global problems. But can we do so in the time available? Exponential growth has of course been with us for a very long time. But for most of recorded history it has been very slow, except for occasional localized spurts. Typical doubling times, for population, GDP, or total energy use, for example, have probably been at least a century and often much more.[1] Now they are of the order of a few decades. Previous local explosions of population or wealth, as in Britain at the time of the industrial revolution, have almost always relied heavily on exploitation of new natural resources and expansion into hitherto unexploited areas—in that case coal and the colonial empire. There are now few places left to expand into, and it is at least doubtful whether we can maintain the required pace of development of new resources.

The shortening time scale is perhaps the most important single point in the whole debate. For it is man's lifetime that fixes the scale for us: we cannot respond by major changes of social structure in less than a generation (except perhaps by revolution). Yet this point is hardly mentioned by Beckerman. Normal economic mechanisms—the feedbacks on which he places so much reliance—take time to operate. When a resource becomes scarce and expensive there is an incentive to look for new sources and substitutes. But to develop the necessary new technology takes time, money and energy. For example, it has taken about twenty-five years for nuclear power generation to reach the stage of providing an appreciable percentage of our total energy consumption, and even this rate of development has been achieved only by taking what many responsible scientists see as unacceptable risks.[2] Again, while the reduction of effluent from new oil refineries, cited by Beckerman,[3] is welcome, what is significant on a global scale is the time it will take for *most* refineries to reach a similar level. It is hard to believe that this can be less than several decades.

It is clearly important, when time is at a premium, to react to possible dangers as early as we can. In view of the novelty of our situation and the complexity of a global system, it is hardly surprising that scientists disagree in their diagnoses of many of our basic problems. But if we insist on waiting until the risk is so obvious as to be incontrovertible before taking action, it may well be too late to avoid catastrophe. Beckerman tells us that he would not be persuaded to invest his life savings in a scheme to break the bank at Monte Carlo by the argument that no one has offered a better one.[4]

[1] A. Ehrlich and P. Ehrlich, op. cit., pp. 6–7.

[2] See, for example, several articles in the *Bulletin of the Atomic Scientists*, Sept. 1971, and also ibid., Nov. 1972, pp. 31–2.

[3] W. Beckerman, op. cit., p. 336.

[4] Ibid., p. 337.

But this is not a fair analogy, for in this situation the alternative of doing nothing at all entails no risk. It may be that many of the dangers now envisaged will turn out to be illusory, but it is overly complacent to ignore all risks which are not yet conclusively proven.

At the very least, it can be argued, a slackening of the pace of economic growth would give us more time to plan ahead It might also give us time to ask what growth is for.[1] It is quite obvious that many things must grow· we need more houses and hospitals; the world desperately needs more food. But it does not follow that all growth is good To treat rising GDP as *a priori* good in itself, as Beckerman appears to do, is to substitute a symbol for the reality. Rising GDP is good if, and to the extent that, it contributes to human well-being and happiness. Of course happiness is hard to measure, and it is much easier for economists to discuss their symbol instead. But much of the 'expansion' represented by rising GDP may actually be illusory, in the form for instance of replacing long-lived products by short-lived ones with no real benefit to anyone It can certainly be argued that a proper evaluation of real human needs would lead to an automatic slowing in the pace of growth

Beckerman states categorically that at 'numerous points where their knowledge overlaps with mine' *The Limits to Growth* team 'have got their facts wrong' [2] He actually quotes four examples, all of which might seem quite damaging on a casual reading. However, at least three of the four are based on fallacious arguments or distortions of what Meadows *et al.* actually wrote

Beckerman first objects to the assertion that the 'few kinds of pollution that have actually been measured have been increasing exponentially'. As we pointed out above, all his counter-examples are local in character. But Meadows *et al* make it quite clear (on the same page from which this quotation is taken)[3] that their statements are intended in a *global* perspective How many pollutants can be shown to be decreasing or static globally?

Secondly, Beckerman accuses Meadows *et al* of 'a flagrant distortion of the data',[4] in a table intended to demonstrate the widening of the gap between rich and poor countries over the period 1961–8 His chief complaint is that the table quotes income levels in 1968 rather than 1961 and that 'naturally, the countries that have grown fastest in any period tend to have higher incomes *at the end of the period*'. At first sight this might seem a very damaging criticism, but in fact it is quite bogus. For, using the same

[1] Several economists have also asked this question See, for example, E J Mishan, *The Costs of Economic Growth* (London Staples Press, 1967)

[2] W Beckerman, op cit, p 339

[3] Meadows *et al.*, op, cit , p 69 See also p 71.

[4] W Beckerman, op cit , p 340 See also Meadows *et al* , op cit , p 42

figures, one can easily verify[1] that the countries that have grown fastest were also those with the higher incomes in 1961.

Thirdly, Beckerman states that 'the old scare about the glasshouse effect of carbon dioxide in the atmosphere is quoted as if it were established fact though everybody knows that this particular scare story has been subjected to very damaging criticism' (by the Royal Commission on the Environment). This is simply not true. The discussion of the effects of increasing carbon dioxide in *The Limits to Growth*[2] is, in fact, very restrained, and has in no way been invalidated The Royal Commission's report does indeed severely criticize the suggestion that one effect might be a large-scale melting of polar ice-caps But that suggestion was not made or even mentioned by Meadows *et al*

Similarly, Beckerman claims[3] that the 'Doomsday model' is refuted by the existence *now* of widespread poverty and starvation and scarcity of materials—on the grounds that it is these facts, rather than the predictions of doom, which have induced the countries in question to take action This is strange logic It may be an unfortunate fact of life that governments are at the moment often persuaded to action only by the threat of immediate crisis, but it is hard to see how this refutes the model It merely highlights the need to take a longer term view

It is as well to realize that the debate is as much about values as about detailed facts and figures, and that the whole of Beckerman's article is informed by a set of values which many scientists would find repugnant[4] For example, we do not accept the proposed antithesis between Mankind and Fishkind[5] even on Beckerman's anthropocentric terms. As Brian Johnson has pointed out,[6] open-air swimming pools for fish (i e clean rivers) and open-air swimming pools for humans can well be one and the same thing More generally Beckerman seems to take a very limited view of human needs, which surely transcend the choice between consumption today and consumption tomorrow.[7]

[1] In fact, the average *per capita* GNP in the five fastest- and slowest growing among the ten most populous countries in 1968 were $1,820 and $100 respectively, while in 1961 they were $1,380 and $95 (according to the same World Bank data) Of course, it may be that GNP is in any case not a good measure to use in this comparison, but that is another question

[2] Op cit, pp. 71–81

[3] W Beckerman, op cit, p 339.

[4] That Beckerman is well aware of the partisan character of his work is shown by a footnote to another recent paper, where he states 'The following few paragraphs present the "official" party line about the role of the economist as the detached adviser on optimal strategies for somebody else's value judgements Personally, I don't subscribe to this doctrine, and I regard the economist as a special kind of propagandist But if this were made widely known our propaganda would be less effective, which is why I make this point in a footnote where nobody is likely to read it' See 'Why we need economic growth', *Lloyd's Bank Review*, no. 102, p. 2 (Oct 1971).

[5] W. Beckerman, op. cit., p. 330.

[6] B. Johnson, 'Keynes' limits to growth revisited', *Teach-in for Survival* (ed M Schwab, London Robinson and Watkins, 1972), pp. 40–1.

[7] See W. Beckerman, op. cit., p 342.

Finally, consider Beckerman's objections to wh
the most fundamental flaw in the whole Doomsday ana
tive character of the model. He offers a model[1] in which the popu
an individual country may 'overshoot the optimum level, leading
collapse of the system through shortages of output in general, or foo
particular, with a resulting eventual collapse through starvation, dis
and so on'. Then he suggests that if these cycles in different countrie
all out of phase, then 'all that happens is that some countries' populat
are rising while others' are falling, and taking the world as a whole it
bumbles along in the usual old way'. The most obvious objection to
model is that it simply does not correspond to what is happening in
real world. At the present time, and for the past quarter century, the
tions of most parts of the world are, and have been, increasing at an un
cented rate[2] A very few are nearly static, but where are the coun
whose populations are on a downward course? In the real world, the c
are not out of phase. Terrible as are the conditions of many people to
they are not nearly bad enough to validate Beckerman's model

It is surely astonishing that someone who claims[3] to be more intere
in human beings than are the natural scientists can contemplate
apparent equanimity the prospect of a world in which country after cou
suffers a catastrophic decline of population through starvation and dis
This is indeed what may happen if we persist on our present course—ex
that, because of the strong coupling provided by global trade, there i
reason to believe that the collapses may not all coincide But this is
'bumbling along in the usual old way' It is stark horror on a world-w
scale If this is their idea of humanity, God save us from hum
economists!

c/o T W B Kibble, Imperial College, London

[1] W Beckerman, op cit, pp 338 9
[2] A Ehrlich and P Ehrlich, op cit, p 6
[3] W Beckerman, op cit, p 330.